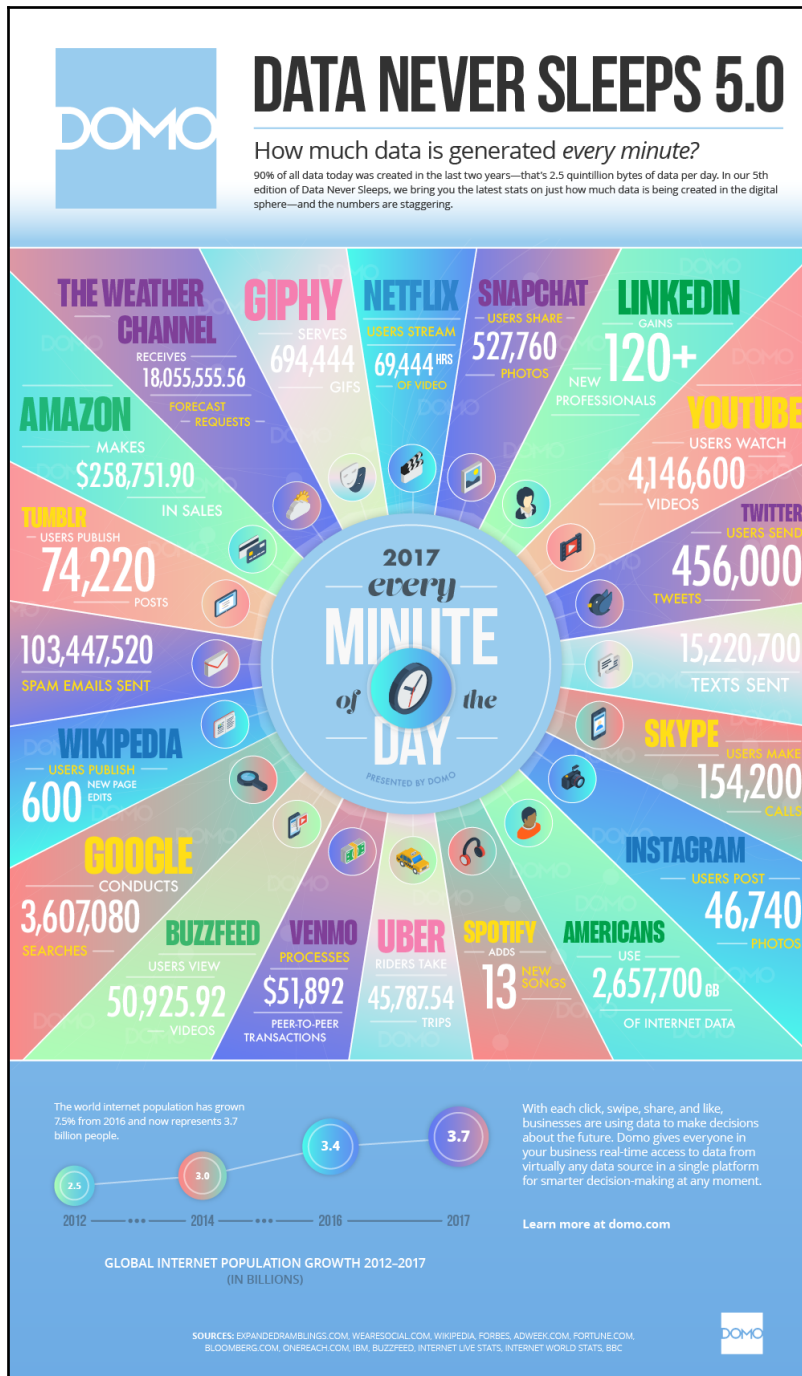
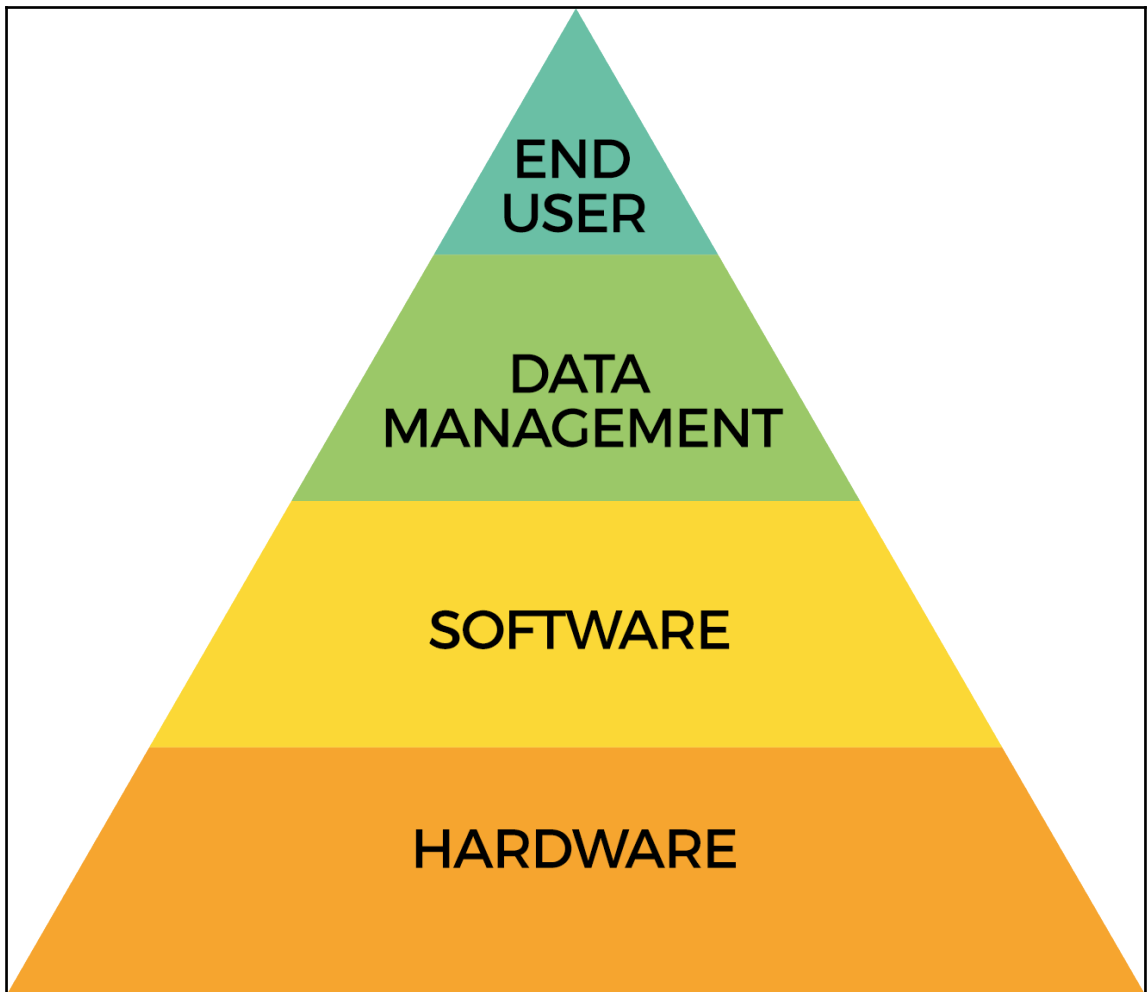


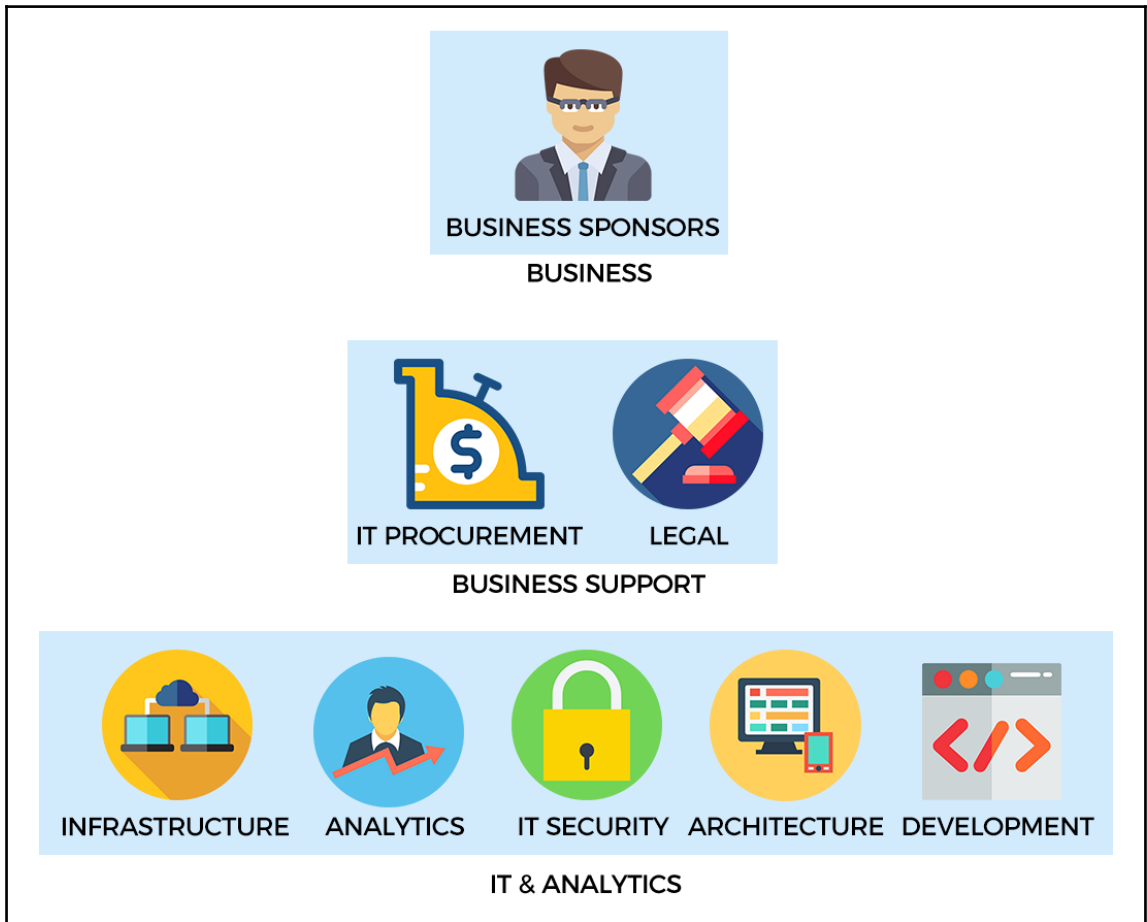
Chapter 1: Too Big or Not Too Big



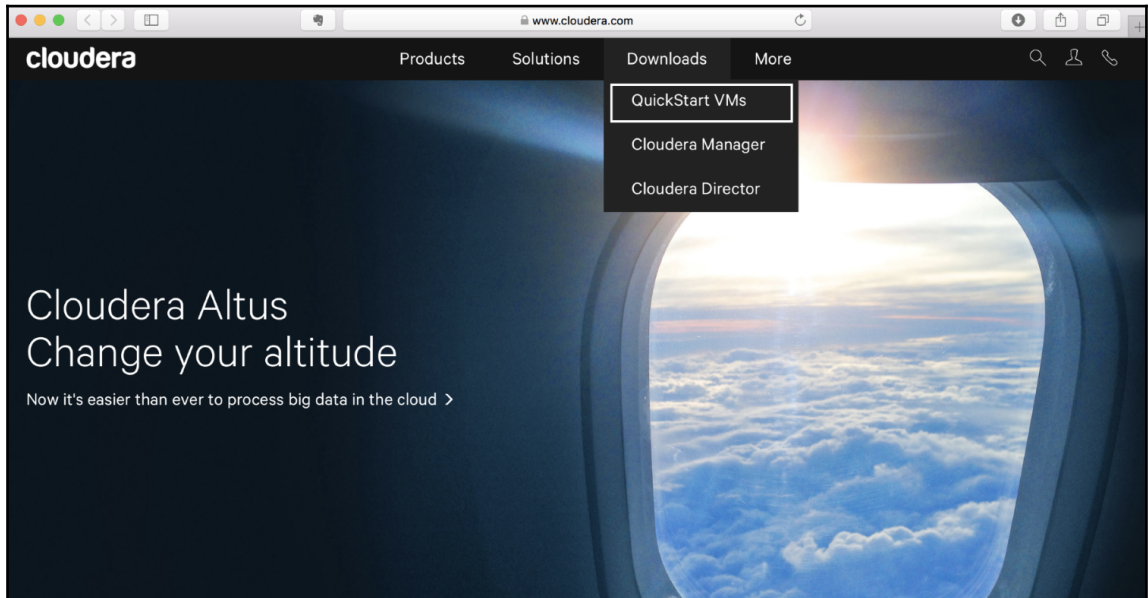


Chapter 2: Big Data Mining for the Masses

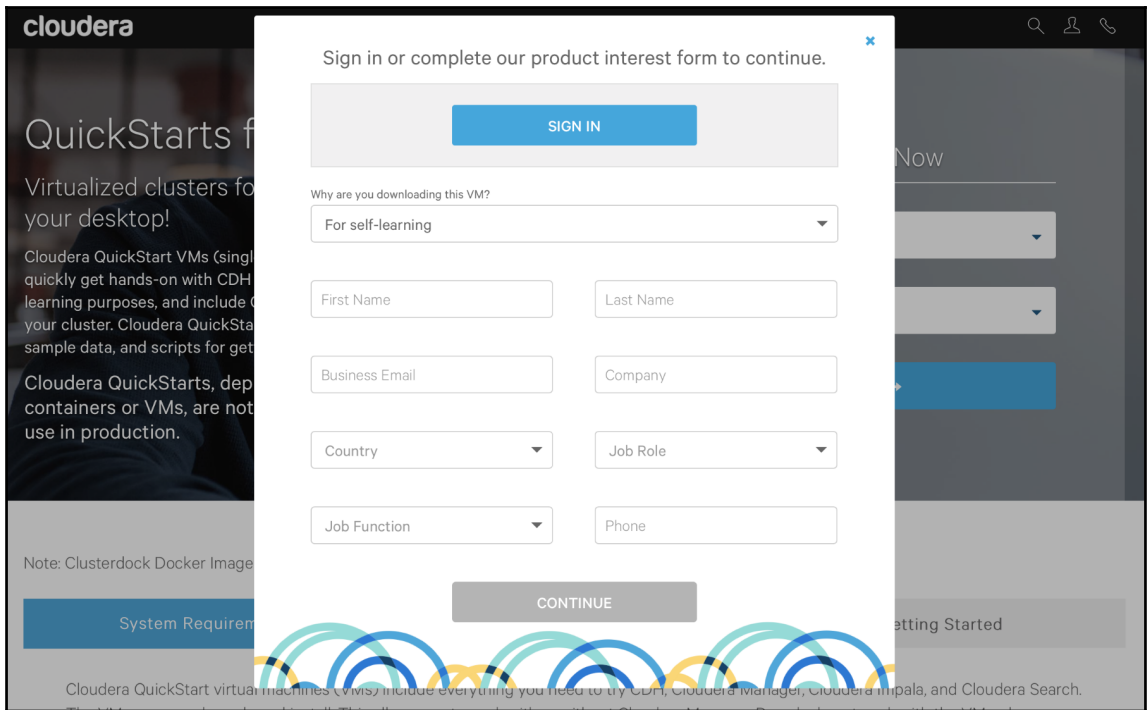
$$\text{E-R Ratio} = \frac{\text{Time} + \text{Cost} + \text{Number of Resources} + \text{Criticality of Use Case}}{\text{Business Value}}$$



Chapter 3: The Analytics Toolkit



The screenshot shows the Cloudera website's 'QuickStarts for CDH 5.10' page. The header includes the Cloudera logo and navigation links for Products, Solutions, Downloads, and More. The main content area features a large image of a man with glasses and a dark sweater. To the right of the image is a 'Get Started Now' section with a 'Version' dropdown menu set to 'QuickStarts for CDH 5.10' and a 'SELECT A PLATFORM' dropdown menu with options for Virtual Box, VMWare, KVM, and Docker Image. Below this is a note about the Docker Image availability for CDH 5.8 and 5.10. At the bottom, there is a navigation bar with three buttons: 'System Requirements' (highlighted in blue), 'Installed Products', and 'Getting Started'.



The screenshot shows the Cloudera QuickStart VM setup interface. A modal dialog box titled "Please Read and Accept our Terms" is centered on the screen. The dialog contains the following text:

Cloudera Standard License and Oracle JDK Notice

Cloudera Standard License

Version 2016-11-08

END USER LICENSE TERMS AND CONDITIONS

THESE TERMS AND CONDITIONS (THESE "TERMS") APPLY TO YOUR USE OF THE PRODUCTS (AS DEFINED BELOW) PROVIDED BY CLOUDERA, INC. ("CLOUDERA").

PLEASE READ THESE TERMS CAREFULLY.

IF YOU ("YOU" OR "CUSTOMER") PLAN TO USE ANY OF THE PRODUCTS ON BEHALF OF A COMPANY OR OTHER ENTITY, YOU REPRESENT THAT YOU ARE THE EMPLOYEE OR AGENT OF SUCH COMPANY (OR OTHER ENTITY) AND YOU HAVE THE AUTHORITY TO ACCEPT ALL OF THE TERMS AND CONDITIONS SET FORTH IN AN ACCEPTED REQUEST (AS DEFINED BELOW) AND THESE TERMS.

I accept the Terms of Cloudera's Standard License and Oracle JDK Notice

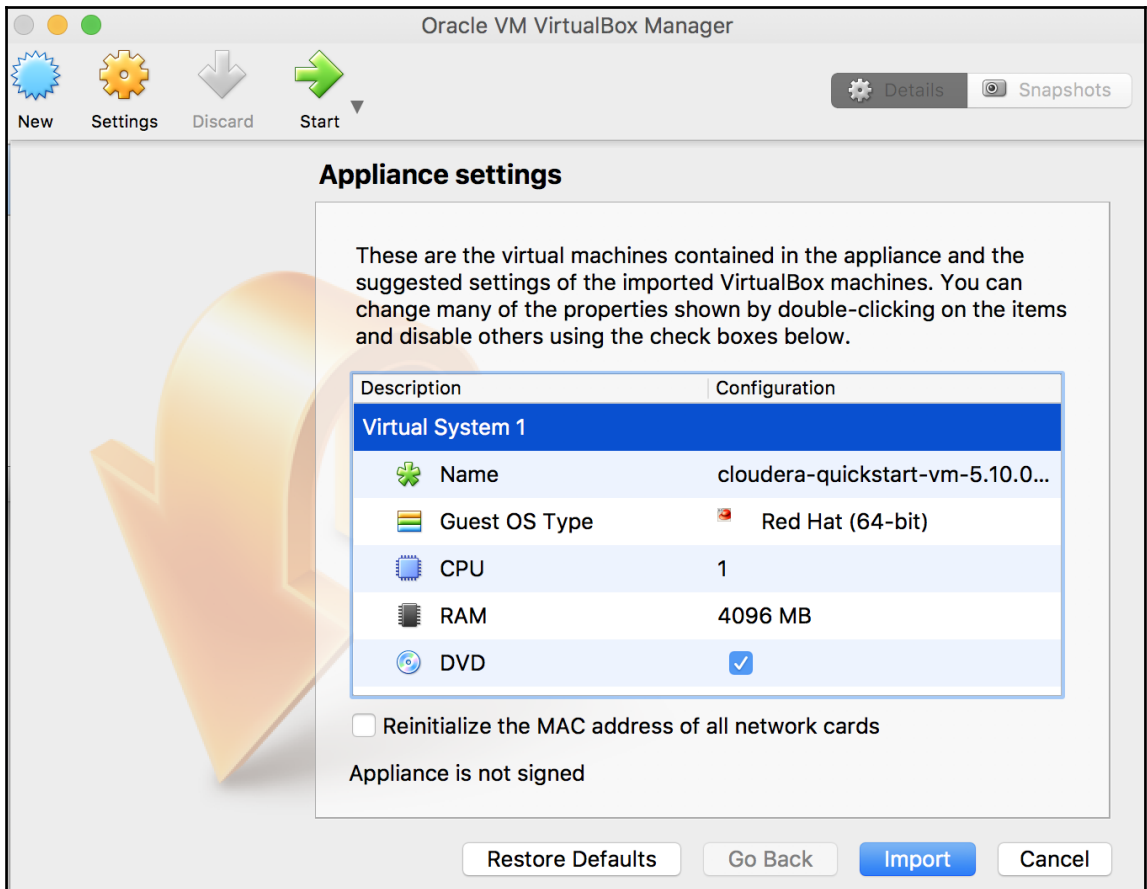
SUBMIT

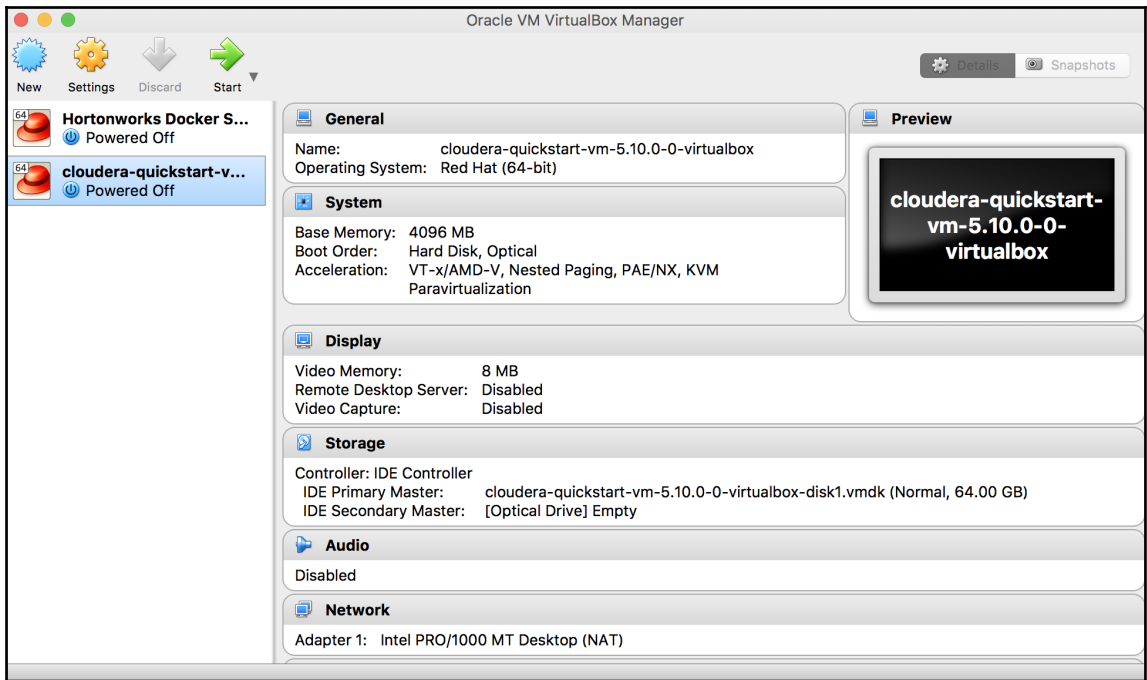
The background interface shows the Cloudera logo and text: "QuickStarts for Virtualized clusters for your desktop!". Below this, it says "Cloudera QuickStart VMs (single) quickly get hands-on with CDH for learning purposes, and include sample data, and scripts for getting started." Another section says "Cloudera QuickStarts, deployed in containers or VMs, are not intended for use in production." At the bottom, there is a note: "Note: Clouderdock Docker Image" and a "System Requirements" button.

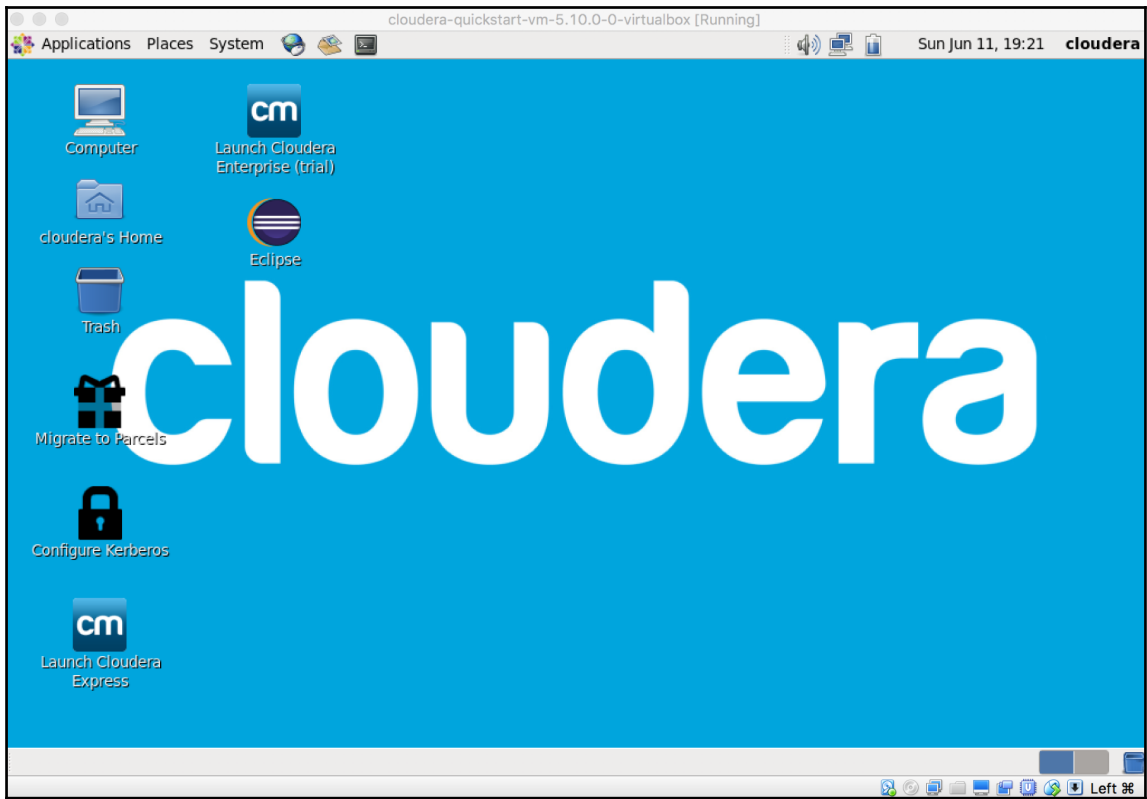
The screenshot shows a download progress bar. The title is "Downloads" with a "Clear" button on the right. The file being downloaded is "cloudera-quickstart-vm-5.10.0-0-virtualbox.zip". The progress bar is partially filled with blue. Below the bar, it says "266.7 MB of 5.28 GB (22.8 MB/sec) — 4 minutes remaining". There are close and search icons on the right side of the progress bar.

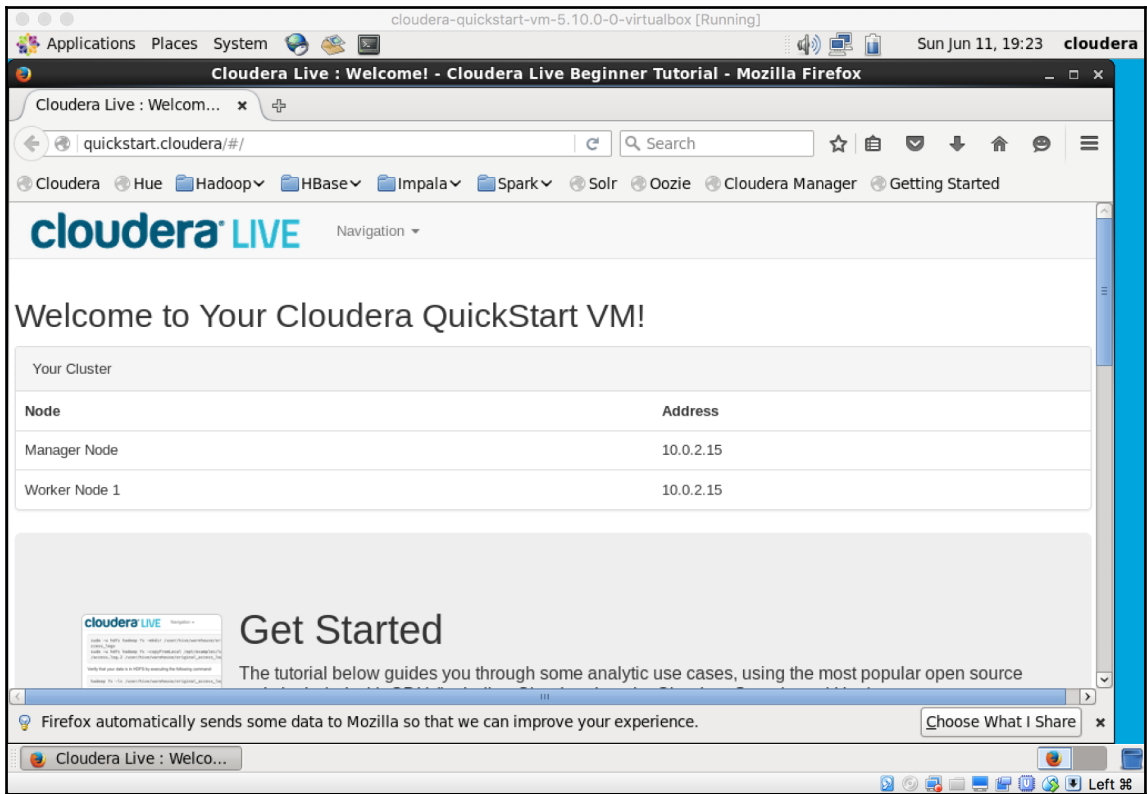
The screenshot shows a file explorer window displaying the contents of a folder named "cloudera-quickstart-vm-5.10.0-0-virtualbox". The files and their sizes are as follows:

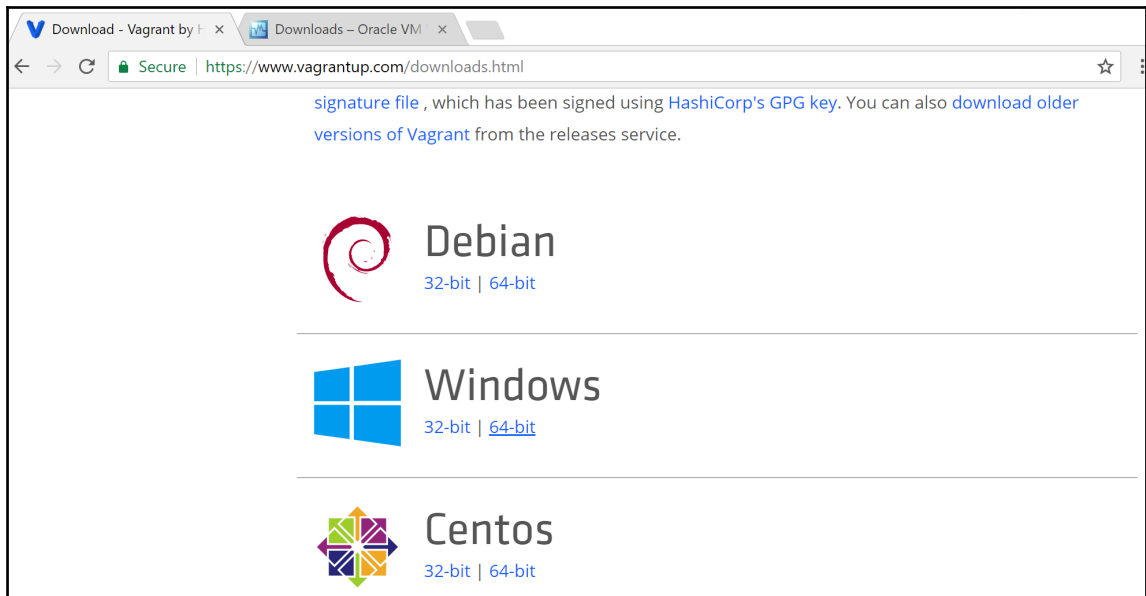
File Name	Size
cloudera-quickstart-vm-5.10.0-0-virtualbox-disk1.vmdk	5.34 GB
cloudera-quickstart-vm-5.10.0-0-virtualbox.ovf	14 KB
cloudera-quickstart-vm-5.10.0-0-virtualbox.zip	5.28 GB

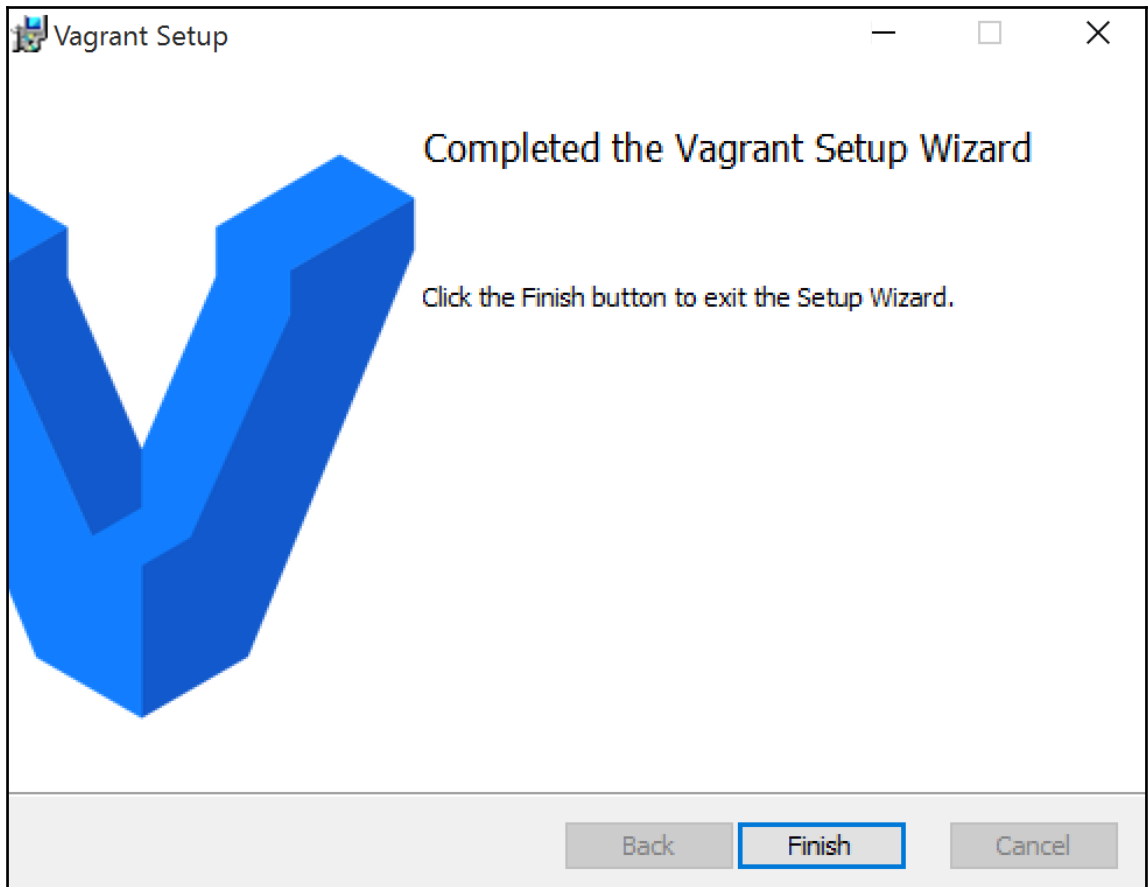


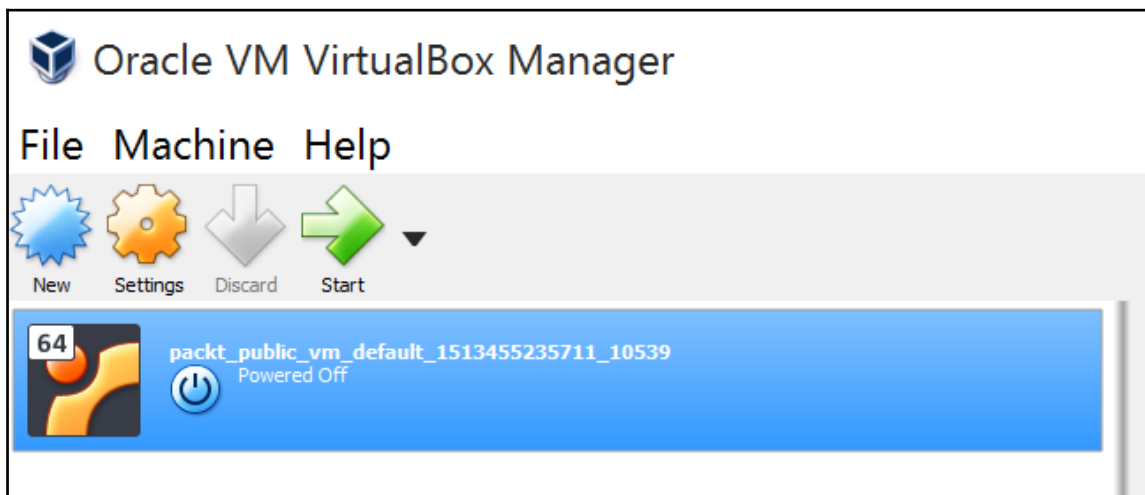
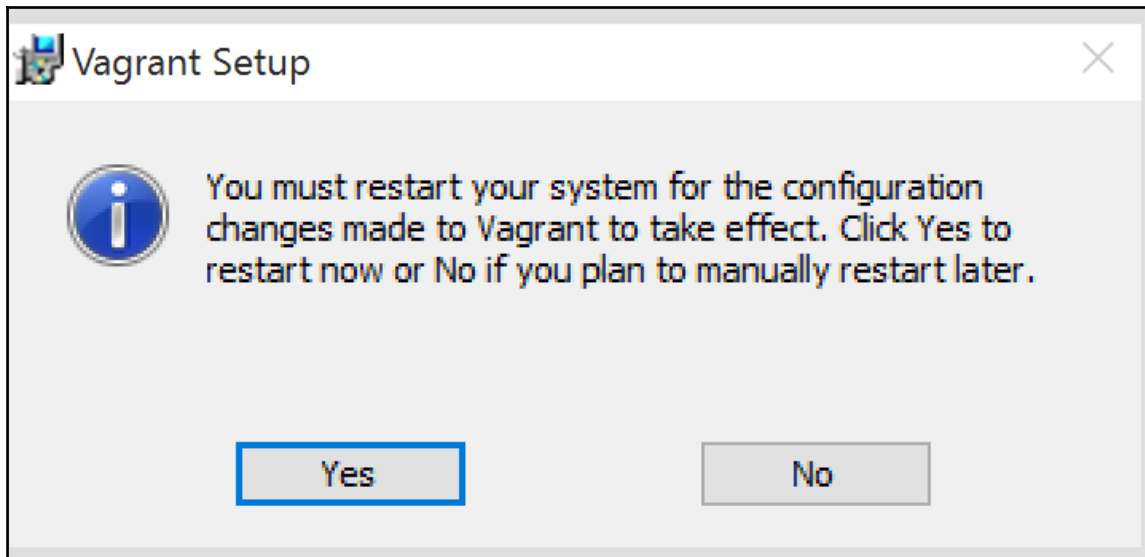












```
Ubuntu 16.04.2 LTS vagrant tty1
vagrant login:
Ubuntu 16.04.2 LTS vagrant tty1
vagrant login: packt
Password:
Last login: Mon Aug 21 16:02:59 UTC 2017 on tty1
Welcome to Ubuntu 16.04.2 LTS (GNU/Linux 4.4.0-83-generic x86_64)

 * Documentation:  https://help.ubuntu.com
 * Management:    https://landscape.canonical.com
 * Support:       https://ubuntu.com/advantage

38 packages can be updated.
6 updates are security updates.

packt@vagrant:~$ _
```



MRAN

[About R](#)
[Microsoft R Open](#)
[Community](#)
[Download](#)

🔍

Microsoft R Application Network

The Microsoft R Portal



R is the world's most powerful programming language for statistical computing, machine learning and graphics as well as a thriving global community of users, developers and [contributors](#).

ANNOUNCEMENT

Microsoft R Open 3.4 is coming soon in May. Download v3.3.3 today.

+ **Microsoft R Open**

Microsoft R Open is the enhanced distribution of open source R from Microsoft Corporation. Enhancements include [multi-core processing](#), a [fixed CRAN repository](#) date, and [reproducible R](#) with the checkpoint package.

📄 **DOWNLOAD NOW**

📅 **CRAN Time Machine**

For the purpose of [reproducibility](#), MRAN hosts [daily snapshots](#) of the CRAN R packages and R releases as far back as Sept. 17, 2014. Use our [Time Machine](#) to browse CRAN contents from the past.

🔍 **BROWSE SNAPSHOTS**

⚙️ **R Packages**

Packages extend R with new functions and data. Whether you're using R to optimize portfolios, analyze genomic sequences, or to predict component failure times, experts in every domain have made resources, applications and code available for free online.

🔍 **EXPLORE PACKAGES**

Microsoft R Open


- [About Microsoft R Open](#)
- [Download Microsoft R Open](#)

Packages

- [Explore Packages](#)
- [Explore Task Views](#)

Connect

- [Contact](#)
- [Licenses](#)



MRAN

[About R](#)
[Microsoft R Open](#)
[Community](#)
[Download](#)

🔍

Download Microsoft R Open 3.3.3, the enhanced R distribution

Microsoft R Open, **the enhanced distribution of R** from Microsoft, is a complete and free open source platform for statistical analysis and data science. Microsoft R Open 3.3.3 is based on (and 100% compatible with) the statistical language, R-3.3.3. It includes additional capabilities for performance, reproducibility and platform support. [Learn more...](#)

Next Release


Microsoft R Open 3.4 is coming soon in May.

Microsoft R Open & MKL Downloads

While the install of MKL, used for multithreaded performance, is **optional**, we recommend both Microsoft R Open & MKL for optimal performance on Windows and Linux. The OS X operating system has Math Libraries by default.

Platforms (64-Bit only)	Downloads
Windows - Windows 7.0 (SP1), 8.1, 10 and Windows Server® 2008 R2 (SP1), 2012 <small>SHA 256: 7f82be9191b9791c1af8e6d17e1f4b493522aebd2d6aad0f0d86a05a74881ec89</small>	↓ R Open / MKL
Ubuntu - 14.04, 16.04 <small>SHA 256: fff62c70f413d555d49707546fed7c281f09fb8a3273a94c6e834992281cb85b</small>	↓ R Open / MKL
Red Hat Enterprise Linux - 6.5, 7.1 <small>SHA 256: fff62c70f413d555d49707546fed7c281f09fb8a3273a94c6e834992281cb85b</small>	↓ R Open / MKL
SUSE Linux Enterprise Server - 11 <small>SHA 256: fff62c70f413d555d49707546fed7c281f09fb8a3273a94c6e834992281cb85b</small>	↓ R Open / MKL
Mac OS X - 10.9, 10.10, 10.11, 10.12 (uses Accelerate Math Libraries) <small>SHA 256: 08f0de900f12912ccc31465c2ace1ca07261cfc8c7a25cc76877bda7f6d644af</small>	↓ R Open
Source Code - The source is optional , and is not required to run R Open.	↓ Source

[Prerequisites & Install Docs](#) |
 [Forum](#) |
 [News](#) |
 [Past Releases](#)



Licensing of R Open and MKL

Microsoft R Open and Revolution R Open are distributed by Microsoft Corporation under the terms of the [GNU General Public License version 2](#).

Use of the Intel Math Kernel Library (Intel MKL) is governed by the terms of the [Microsoft R Services MKL End User License Agreement](#).



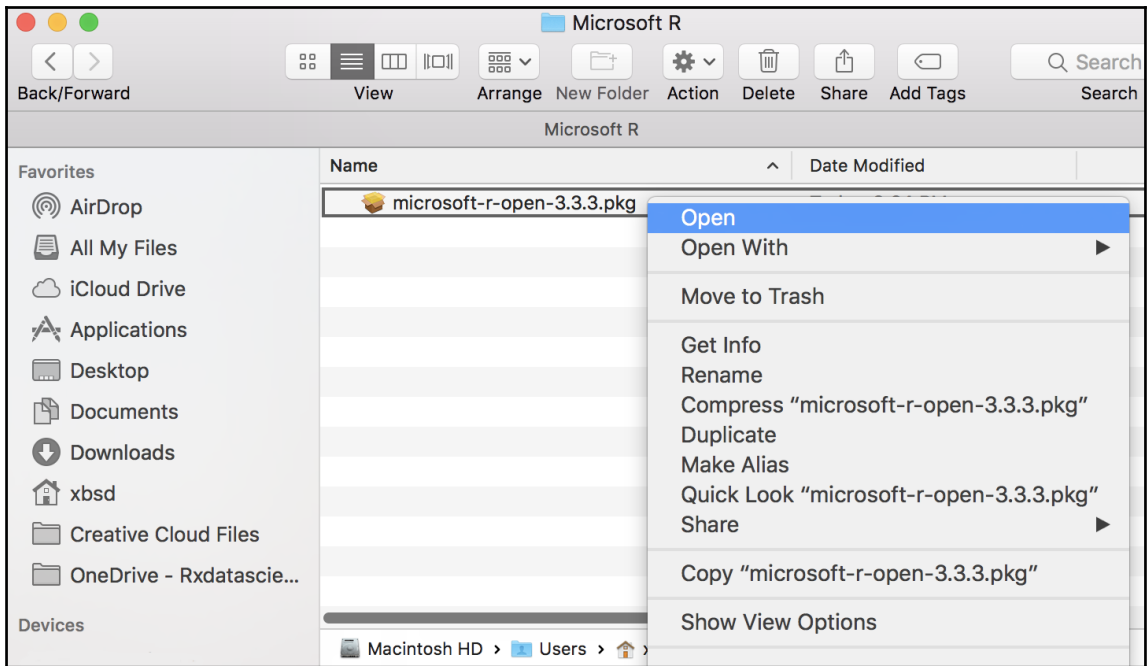
"microsoft-r-open-3.3.3.pkg" can't be opened because it is from an unidentified developer.

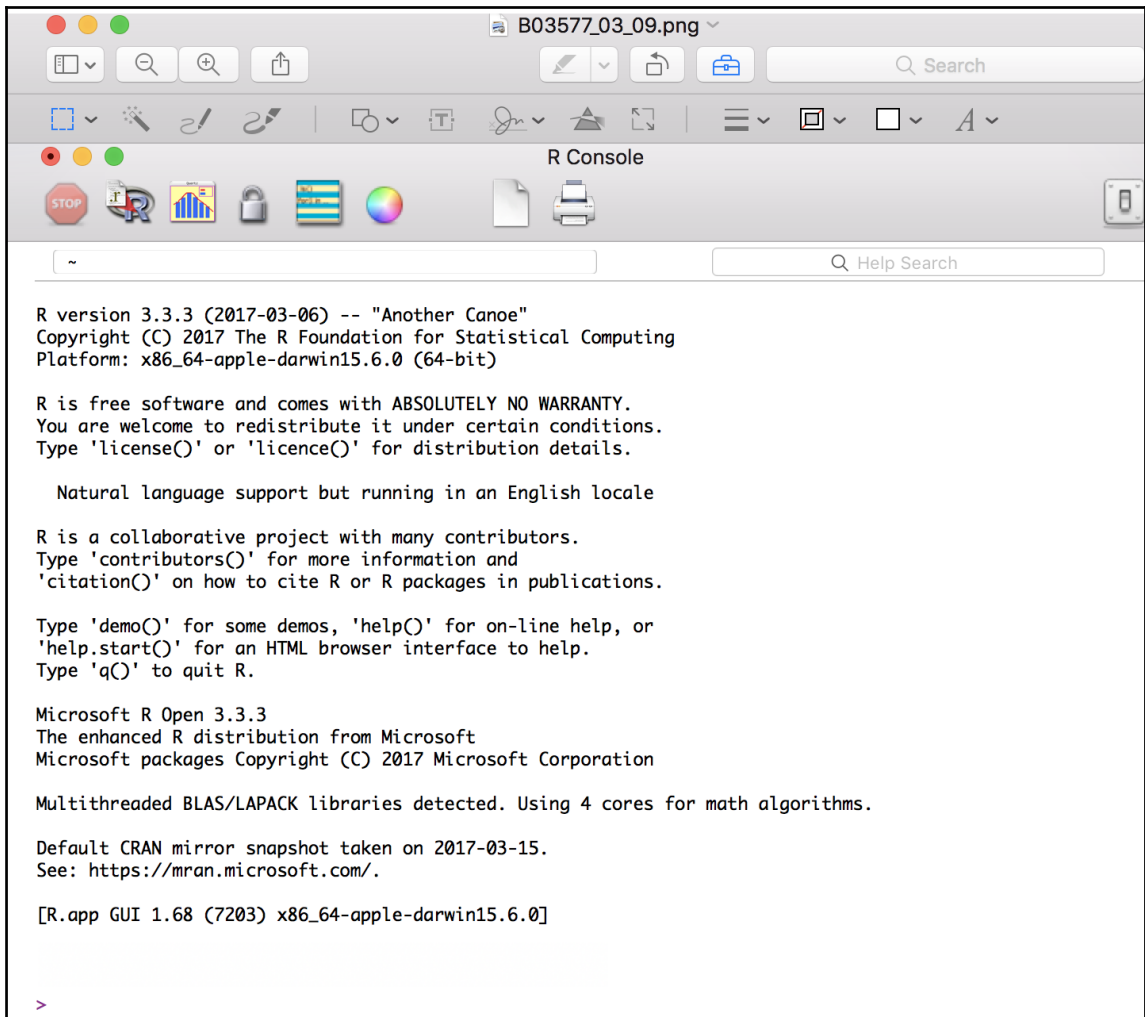
Your security preferences allow installation of only apps from the App Store and identified developers.

Safari downloaded this file today at 9:24 PM from **mran.microsoft.com**.

?

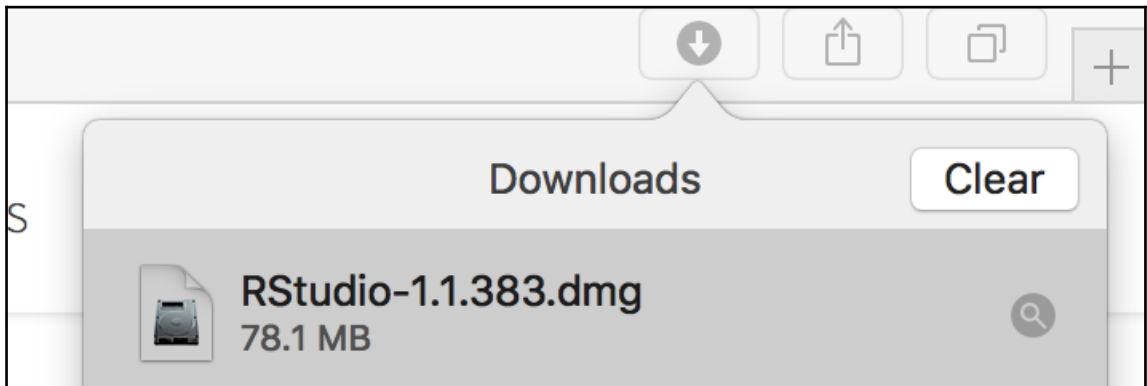
OK

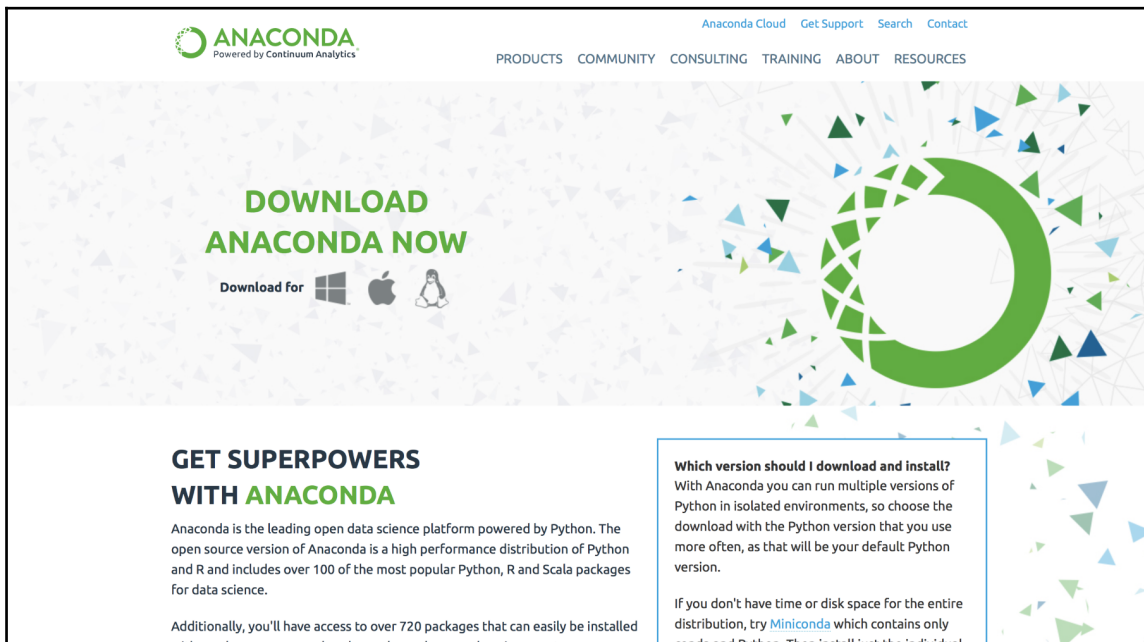


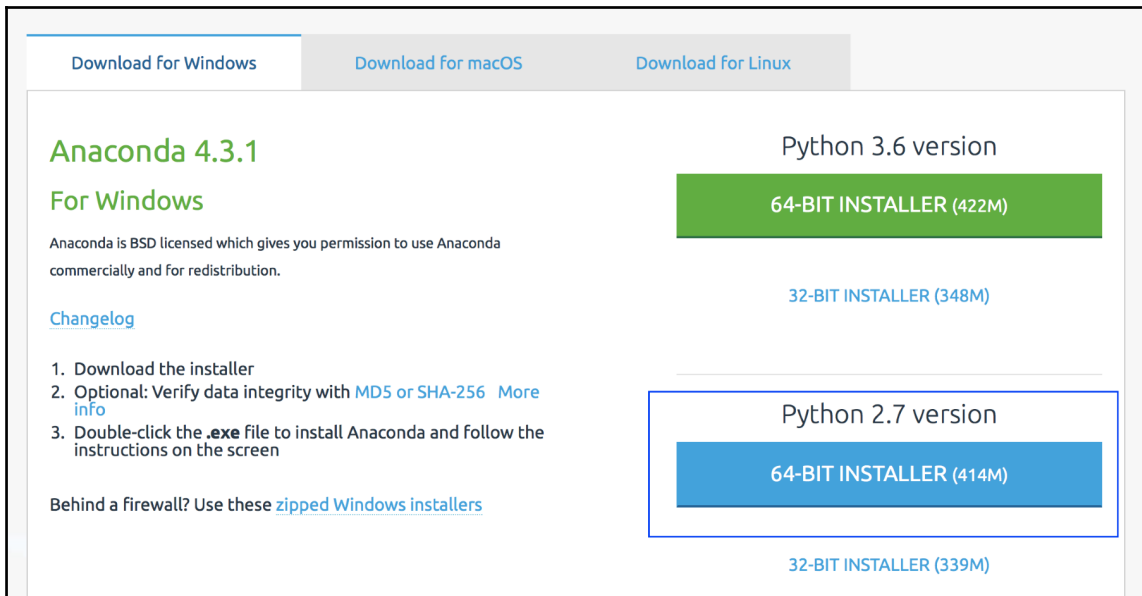


The screenshot shows the RStudio website's pricing page. At the top, the RStudio logo is on the left, and navigation links for 'rstudio::conf', 'Products', 'Resources', 'Pricing', 'About Us', and 'Blogs' are on the right. Below the navigation is a section titled 'Choose Your Version of RStudio' with a sub-header 'Choose Your Version of RStudio'. A paragraph describes RStudio as a set of integrated tools for R. Below this is a table of five product versions with their respective prices and download/buy buttons. A 'DOWNLOAD' button for the 'RStudio Desktop Open Source License' is highlighted with a blue box. Below the table are three rows of features: 'Integrated Tools for R', 'Priority Support', and 'Access via Web Browser', each with green dots indicating availability for the different versions.

	RStudio Desktop Open Source License	RStudio Desktop Commercial License	RStudio Server Open Source License	RStudio Server Pro Commercial License	RStudio Server Pro + RStudio Connect Commercial License
	FREE	\$995 per year	FREE	\$9,995 per year	\$29,995 per year
	DOWNLOAD Learn More	BUY Learn More	DOWNLOAD Learn More	DOWNLOAD Learn More	TALK Learn More
Integrated Tools for R	●	●	●	●	●
Priority Support		●		●	●
Access via Web Browser			●	●	●







Download for Windows Download for macOS Download for Linux

Anaconda 4.3.1

For Windows

Anaconda is BSD licensed which gives you permission to use Anaconda commercially and for redistribution.

[Changelog](#)

1. Download the installer
2. Optional: Verify data integrity with [MD5](#) or [SHA-256](#) [More info](#)
3. Double-click the `.exe` file to install Anaconda and follow the instructions on the screen

Behind a firewall? Use these [zipped Windows installers](#)

Python 3.6 version

64-BIT INSTALLER (422M)

32-BIT INSTALLER (348M)


Python 2.7 version

64-BIT INSTALLER (414M)

32-BIT INSTALLER (339M)

```
MacBook-Pro-3:~ xbsd$ python
Python 2.7.13 |Anaconda 4.3.0 (x86_64)| (default, Dec 20 2016, 23:05:08)
[GCC 4.2.1 Compatible Apple LLVM 6.0 (clang-600.0.57)] on darwin
Type "help", "copyright", "credits" or "license" for more information.
Anaconda is brought to you by Continuum Analytics.
Please check out: http://continuum.io/thanks and https://anaconda.org
>>> █
```

Chapter 4: Big Data With Hadoop

 Nutch / NUTCH-193
move NDFS and MapReduce to a separate project

Agile Board

Description

The NDFS and MapReduce code should move from Nutch to a new Lucene sub-project named Hadoop.

My plan is to do this as follows:

1. Move all code in the following packages from Nutch to Hadoop:
org.apache.nutch.fs
org.apache.nutch.io
org.apache.nutch.ipc
org.apache.nutch.mapred
org.apache.nutch.ndfs

These packages will all be renamed to org.apache.hadoop, and Nutch code will be updated to reflect this.


2. Move selected classes from Nutch to Hadoop, as follows:
org.apache.nutch.util.NutchConf -> org.apache.hadoop.conf.Configuration
org.apache.nutch.util.NutchConfigurable -> org.apache.hadoop.Configurable
org.apache.nutch.util.NutchConfigured -> org.apache.hadoop.Configured

org.apache.nutch.util.Progress -> org.apache.hadoop.util.Progress
org.apache.nutch.util.LogFormatter -> org.apache.hadoop.util.LogFormatter
org.apache.nutch.util.Daemon -> org.apache.hadoop.util.Daemon

3. Add a jar containing all of the above the Nutch's lib directory.


Does this plan sound reasonable?

Issue Links

incorporates  HADOOP-4 initial import of code from Nutch ↑ CLOSED

Activity

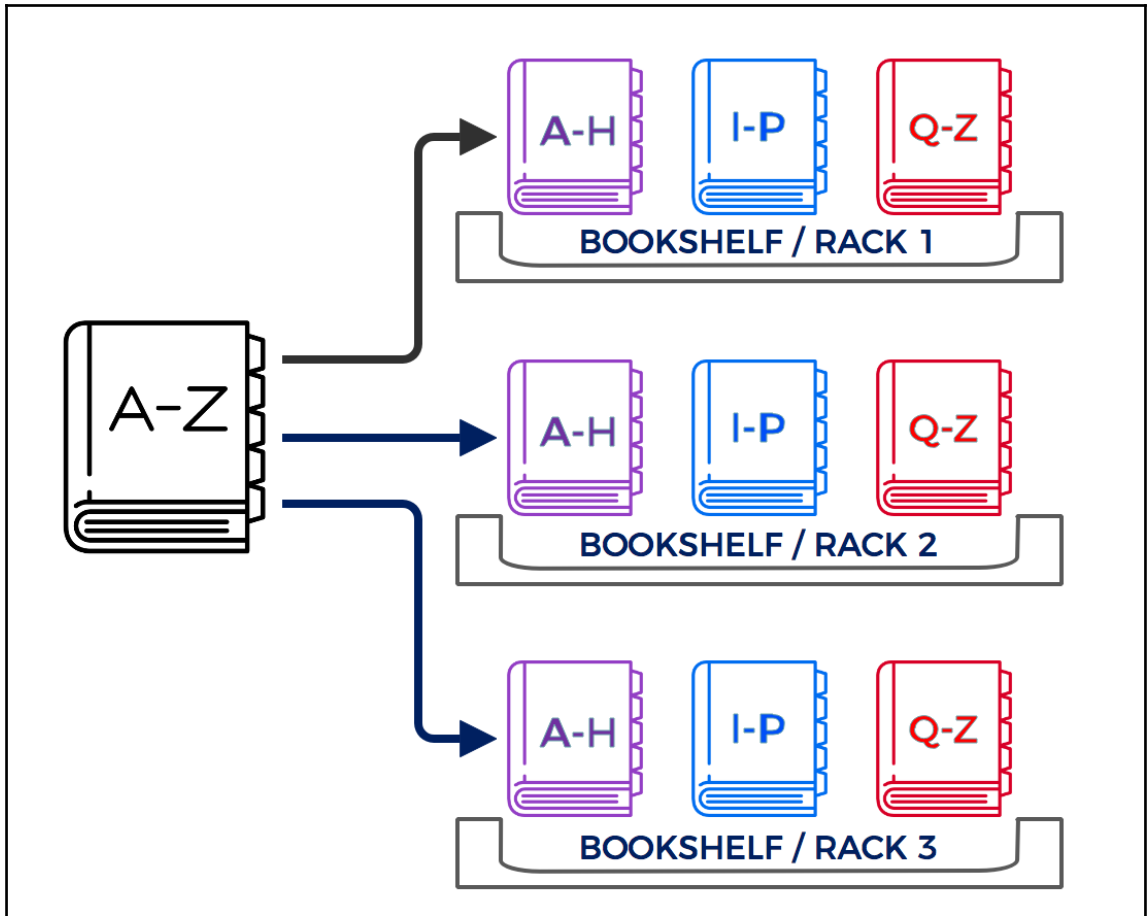
All Comments Work Log History Activity Transitions ↑

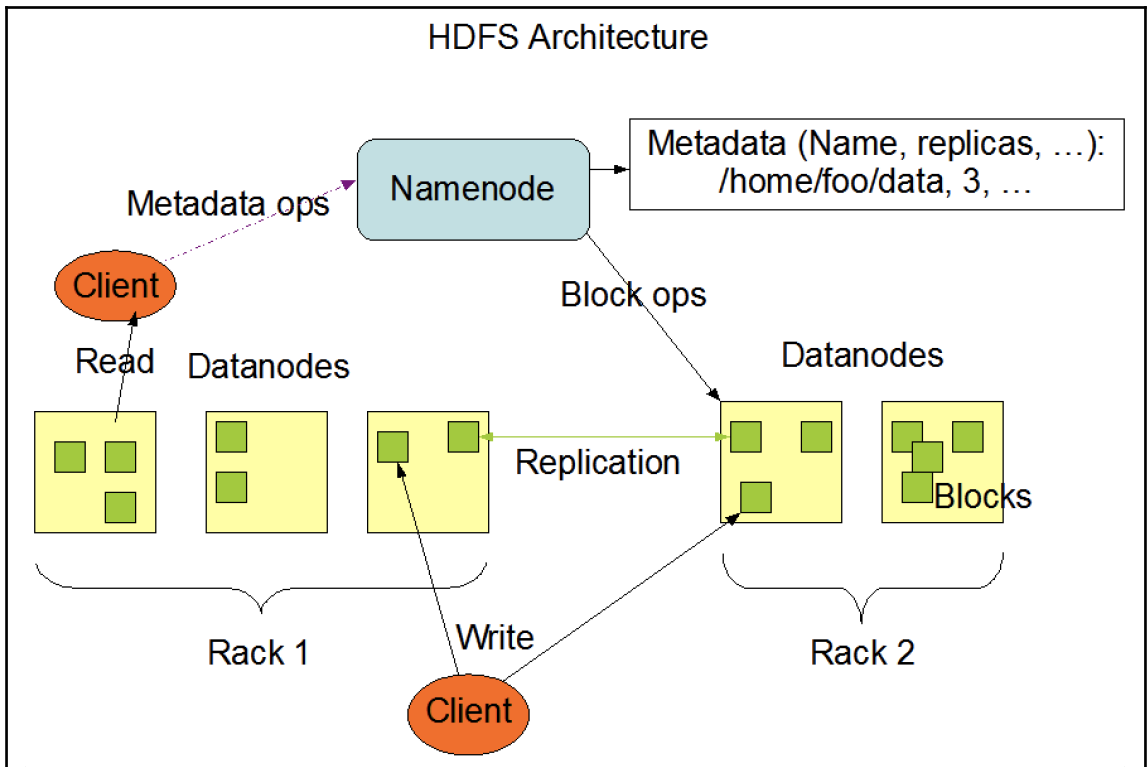
 Doug Cutting created issue - 01/Feb/06 02:52

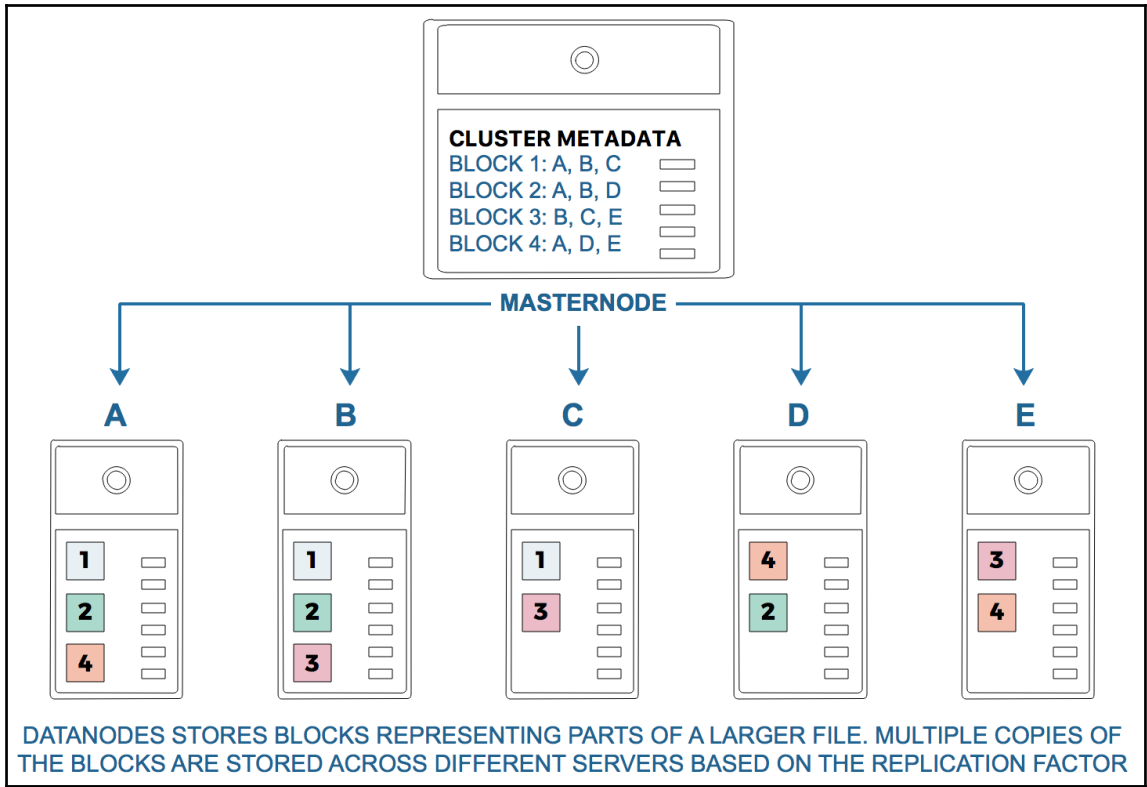
Created: 01/Feb/06 02:52
Updated: 08/Jun/11 21:32
Resolved: 04/Feb/06 09:49

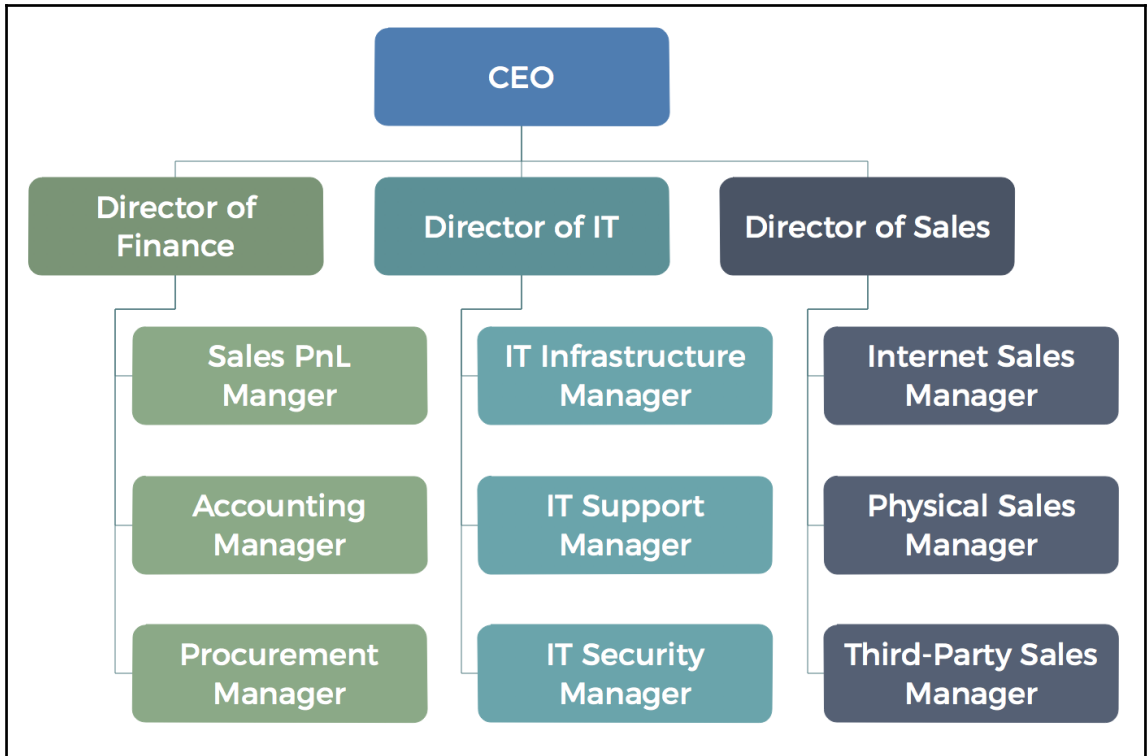
Development
1 commit

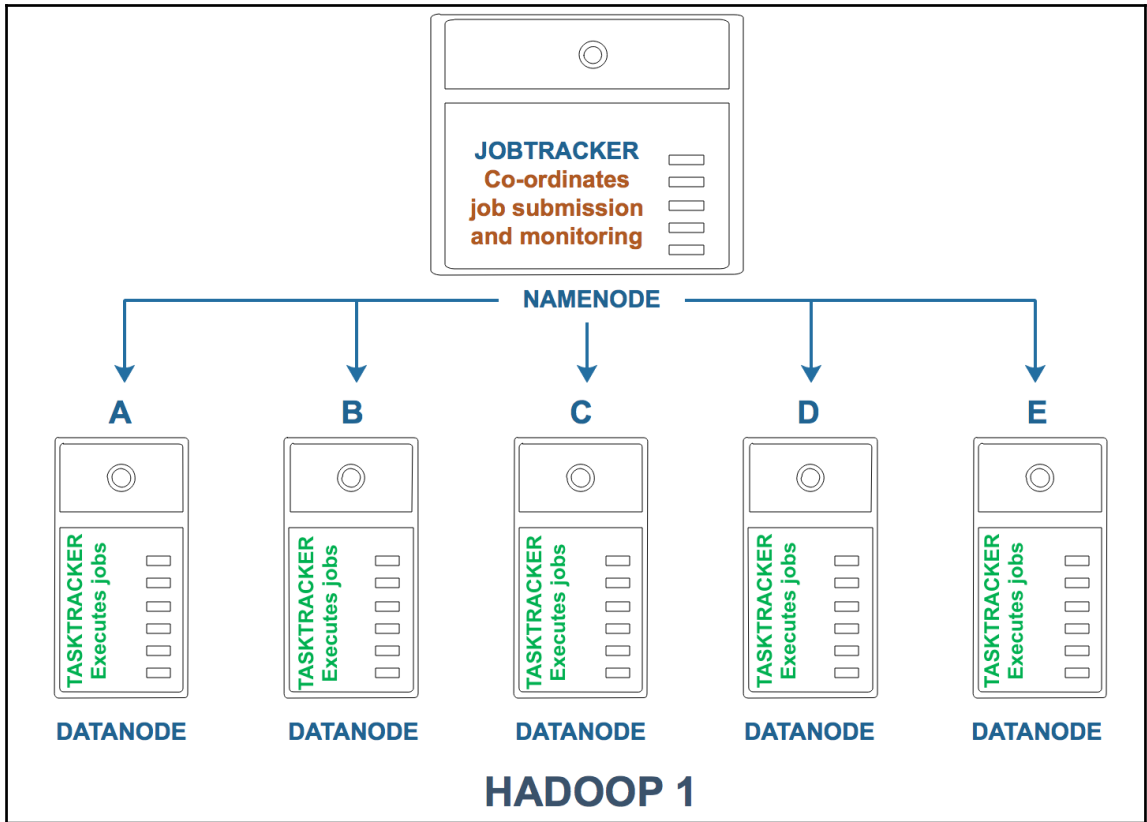
Agile
[View on Board](#)

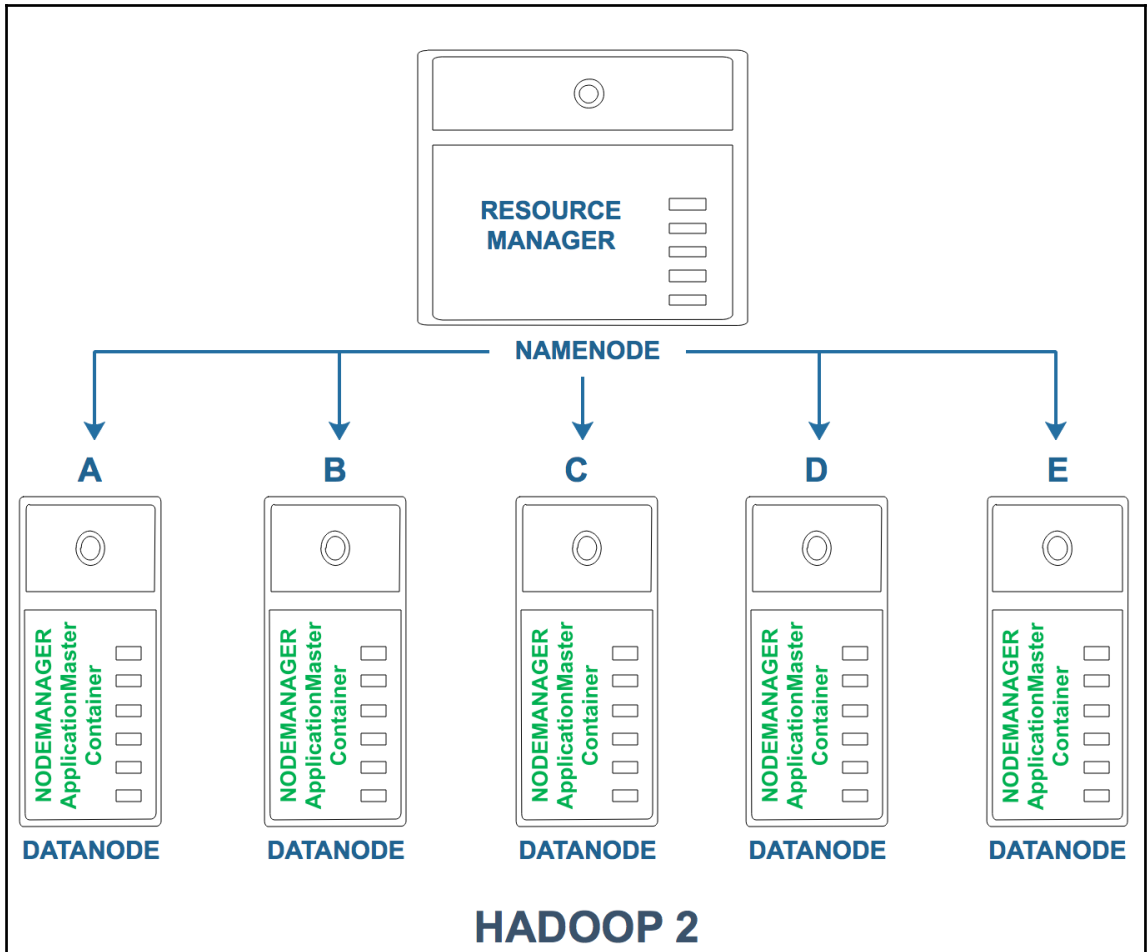


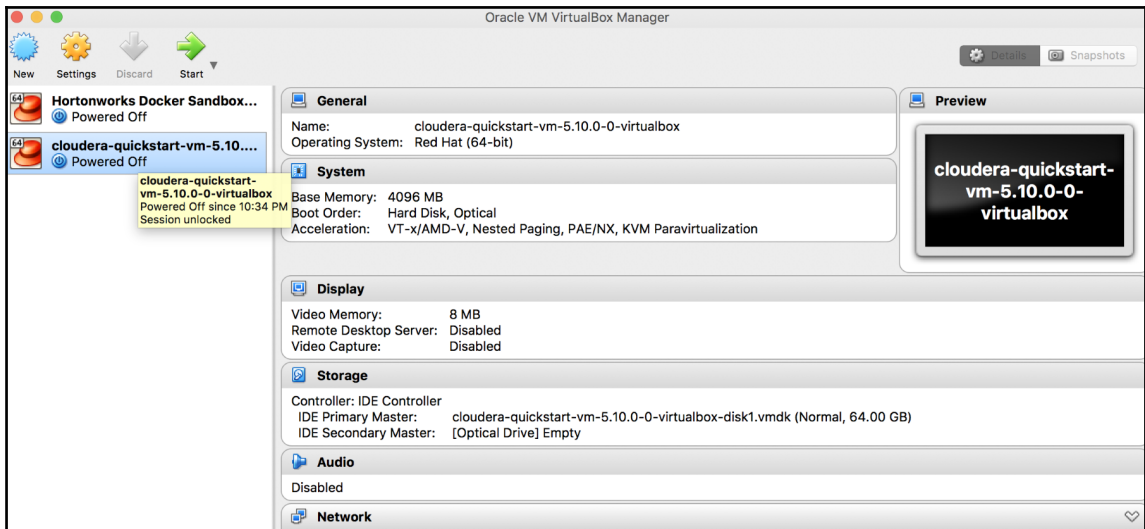
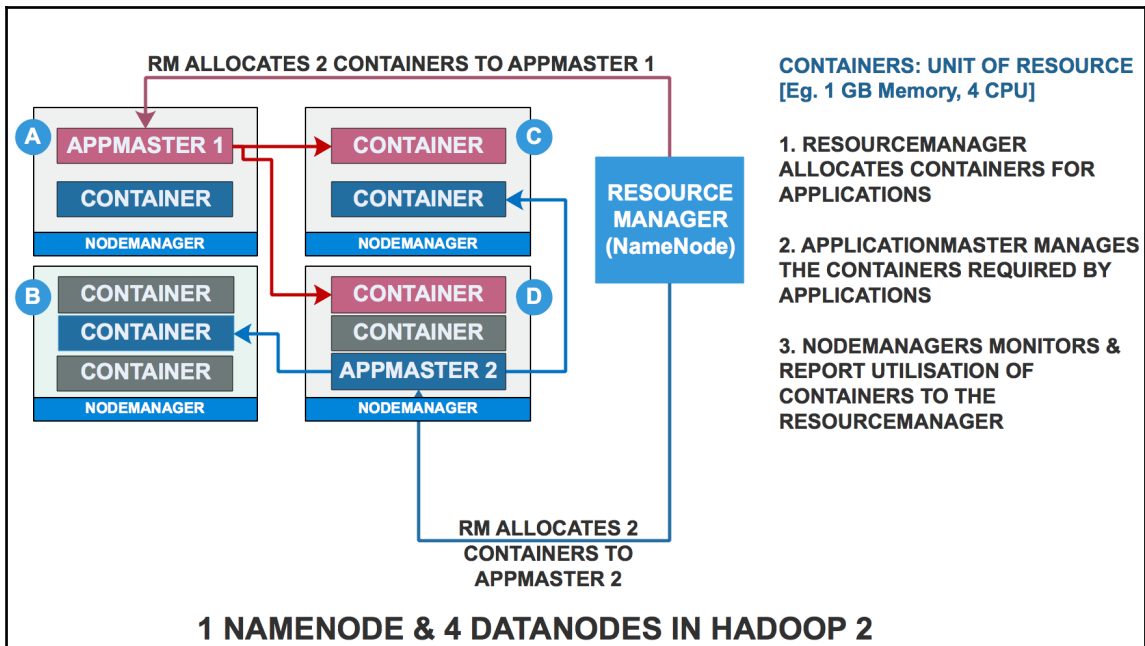


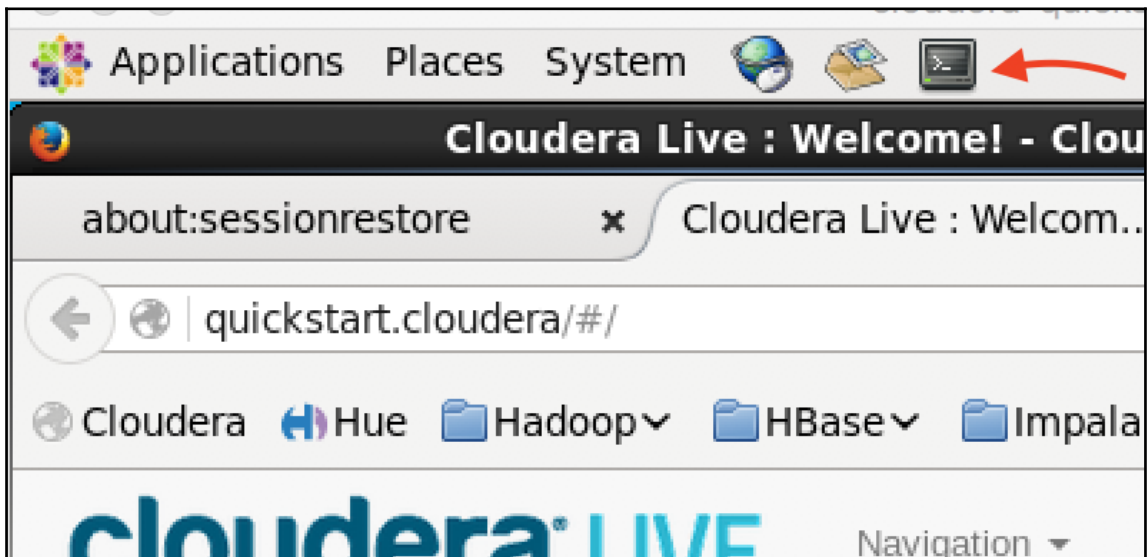
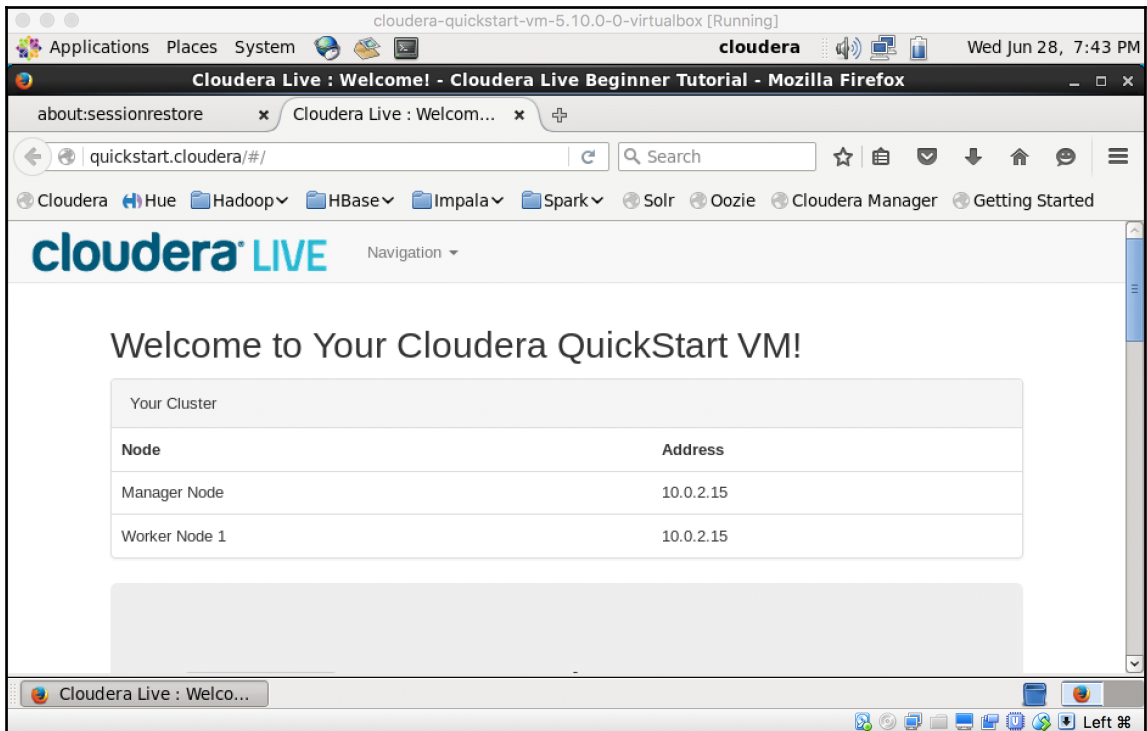


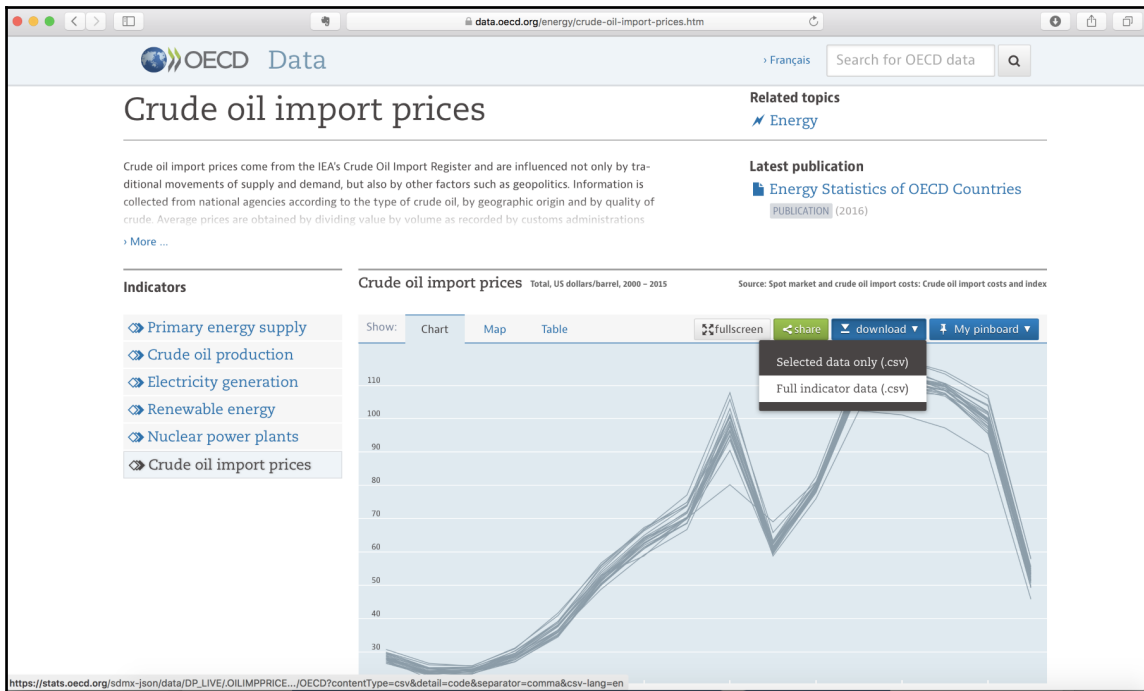








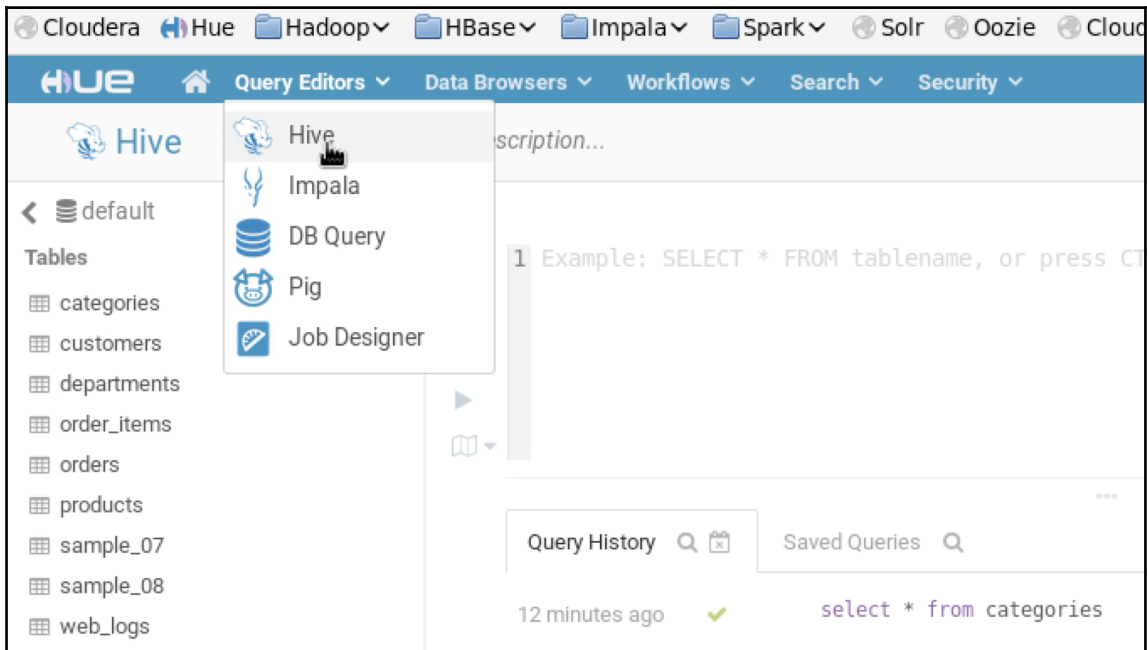
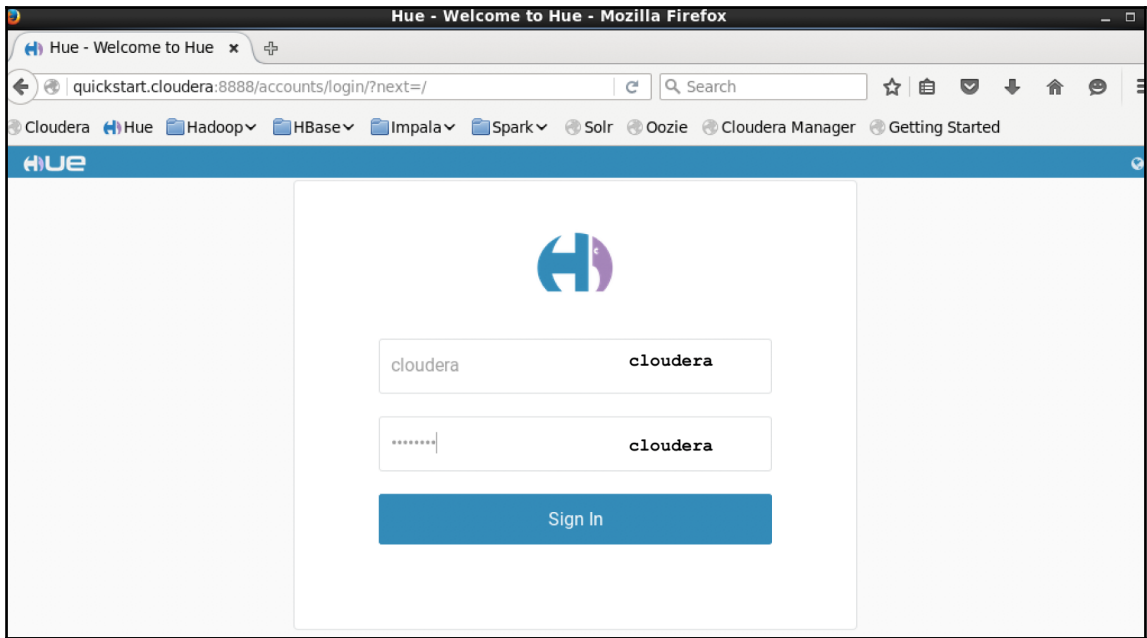




```
[cloudera@quickstart ~]$ cd /home/cloudera;
[cloudera@quickstart ~]$
[cloudera@quickstart ~]$ wget -O oil.csv "https://stats.oecd.org/sdmx-json/data/DP_LIVE/.OILIMPPRICE.../OECD?
contentType=csv&detail=code&separator=comma&csv-lang=en"
--2017-07-22 13:10:40-- https://stats.oecd.org/sdmx-json/data/DP_LIVE/.OILIMPPRICE.../OECD?contentType=csv&
etail=code&separator=comma&csv-lang=en
Resolving stats.oecd.org... 78.41.128.136
Connecting to stats.oecd.org|78.41.128.136|:443... connected.
HTTP request sent, awaiting response... 200 OK
Length: unspecified [text/csv]
Saving to: "oil.csv"

[          <=>          ] 988,528      329K/s   in 2.9s

2017-07-22 13:10:44 (329 KB/s) - "oil.csv" saved [988528]
```



The screenshot shows the Hive Query Editor interface. The left sidebar lists various tables, including 'oil'. The main editor contains the following SQL code:

```

1 CREATE TABLE IF NOT EXISTS OIL
2 (location String, indicator String, subject String, measure String,
3 frequency String, time String, value Float, flagCode String)
4 ROW FORMAT DELIMITED
5 FIELDS TERMINATED BY ','
6 LINES TERMINATED BY '\n'
7 STORED AS TEXTFILE
8 tblproperties("skip.header.line.count"="1");
9
10 LOAD DATA LOCAL INPATH '/home/cloudera/oil.csv' INTO TABLE OIL;
11 SELECT * FROM OIL;
    
```

The results table below the query shows 5 rows of data:

	oil.location	oil.indicator	oil.subject	oil.measure	oil.frequency	oil.time	oil.value
1	"AUS"	"OILIMPPRICE"	"TOT"	"USD_BAR"	"A"	"1980"	31.809999461
2	"AUS"	"OILIMPPRICE"	"TOT"	"USD_BAR"	"A"	"1981"	35.880001061
3	"AUS"	"OILIMPPRICE"	"TOT"	"USD_BAR"	"A"	"1982"	35.419998161
4	"AUS"	"OILIMPPRICE"	"TOT"	"USD_BAR"	"A"	"1983"	30.879999161
5	"AUS"	"OILIMPPRICE"	"TOT"	"USD_BAR"	"A"	"1984"	29.190000531

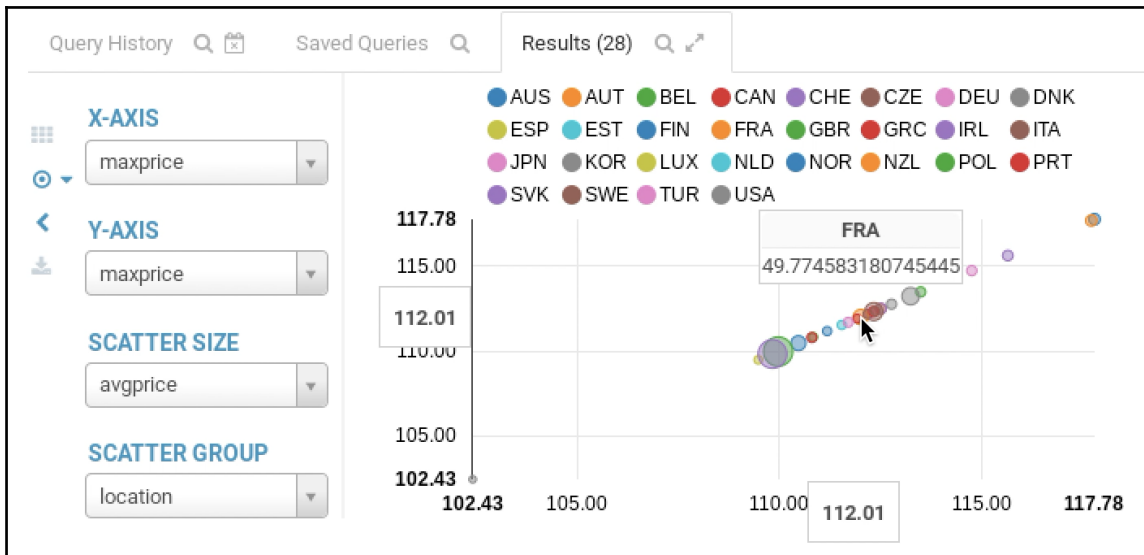
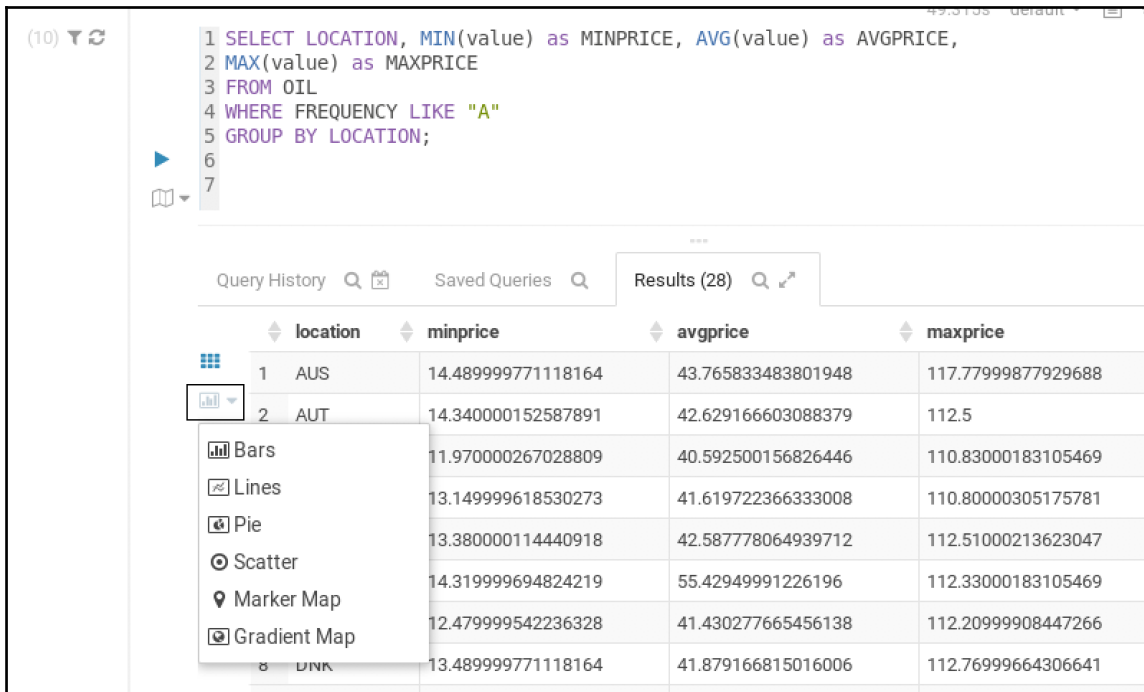
The screenshot shows the Hive Query Editor interface. The main editor contains the following SQL code:

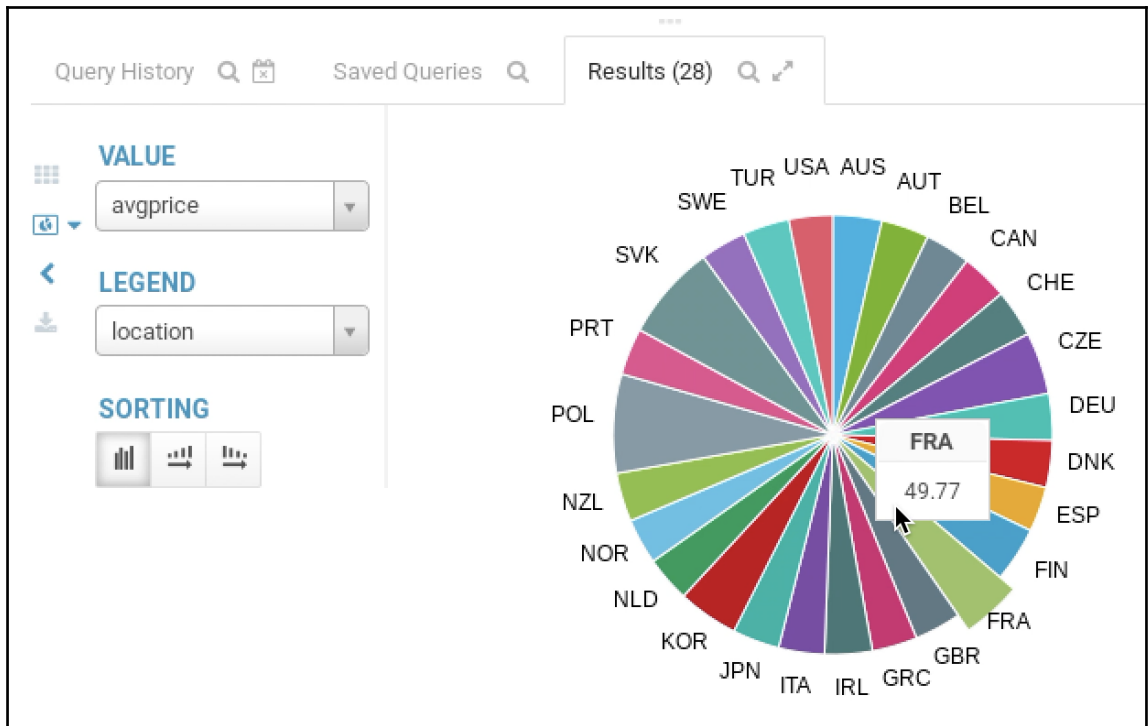
```

1 SELECT LOCATION, MIN(value) as MINPRICE, AVG(value) as AVGPRI,
2 MAX(value) as MAXPRICE
3 FROM OIL
4 WHERE FREQUENCY LIKE "A"
5 GROUP BY LOCATION;
6
7
    
```

The results table below the query shows 4 rows of aggregated data:

	location	minprice	avgprice	maxprice
1	AUS	14.489999771118164	43.765833483801948	117.77999877929688
2	AUT	14.340000152587891	42.629166603088379	112.5
3	BEL	11.970000267028809	40.592500156826446	110.83000183105469
4	CAN	13.149999618530273	41.619722366333008	110.80000305175781





```

1 CREATE TABLE IF NOT EXISTS LATLONG
2   (country String, alpha2 String, alpha3 String, numCode Int, latitude Float, longitude Float)
3   ROW FORMAT DELIMITED
4   FIELDS TERMINATED BY ','
5   LINES TERMINATED BY '\n'
6   STORED AS TEXTFILE
7   TBLPROPERTIES("skip.header.line.count"="1");
8
9
10 LOAD DATA LOCAL INPATH '/home/cloudera/latlong.csv' INTO TABLE LATLONG;
11

```

Success.

Query History

- a few seconds ago ✓ LOAD DATA LOCAL INPATH '/home/cloudera/latlong.csv' INTO TABLE LATLONG
- a few seconds ago ✓ CREATE TABLE IF NOT EXISTS LATLONG (country String, alpha2 String, alpha3 String, numCode Int, latitude Float, longitude Float) ROW FORMAT DELIMITED FIELDS TERMINATED BY ',' LINES TERMINATED BY '\n' STORED AS TEXTFILE TBLPROPERTIES("skip.header.line.count"="1")

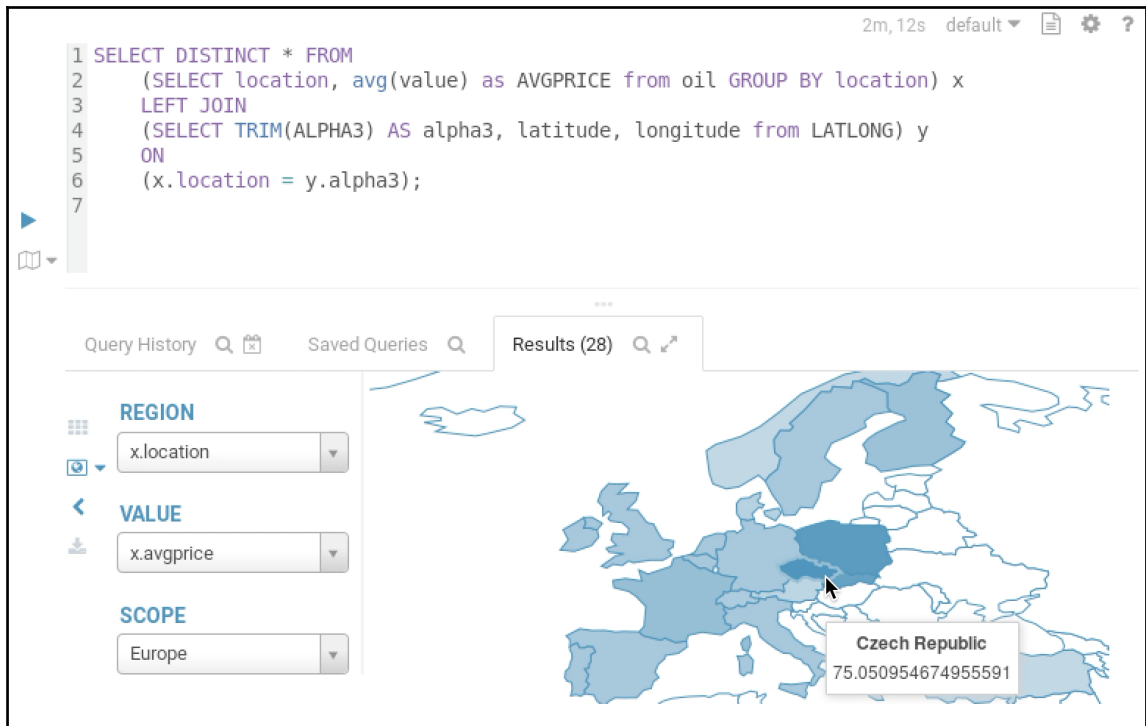
```

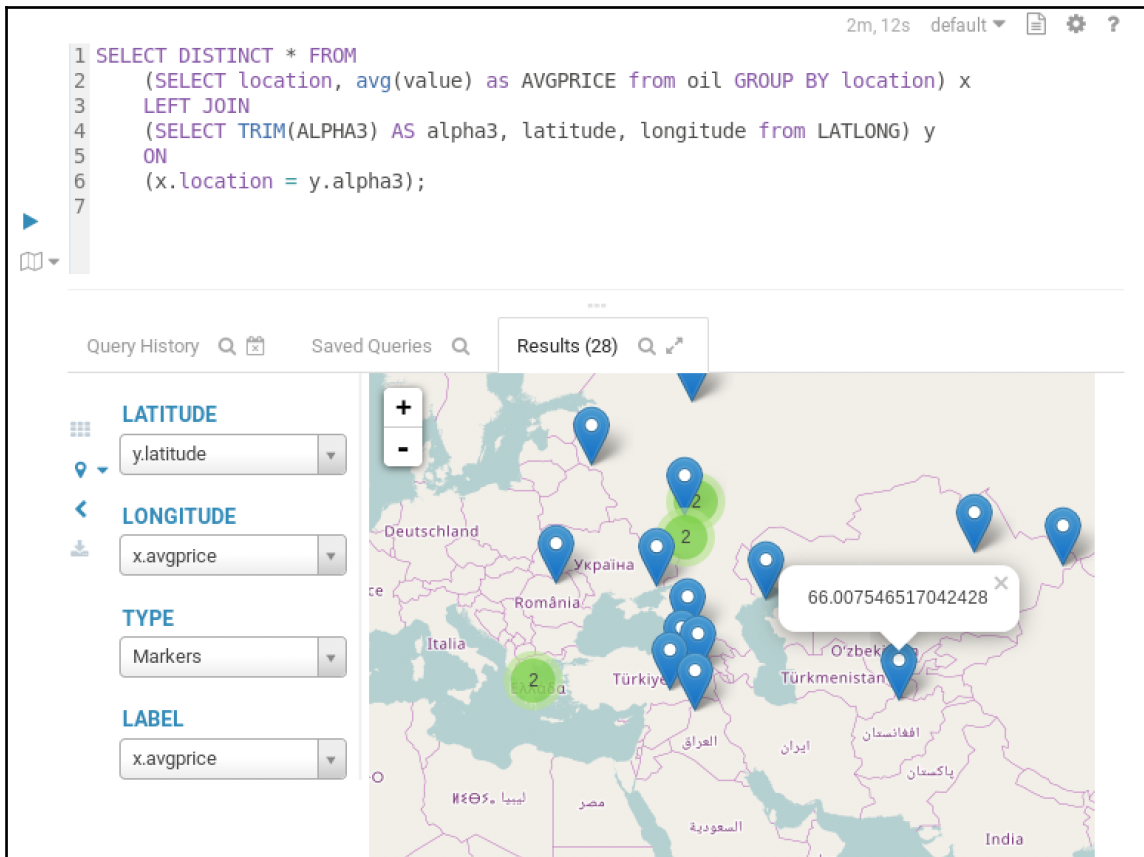
1 SELECT DISTINCT * FROM
2   (SELECT location, avg(value) as AVGPRICE from oil GROUP BY location) x
3   LEFT JOIN
4   (SELECT TRIM(ALPHA3) AS alpha3, latitude, longitude from LATLONG) y
5   ON
6   (x.location = y.alpha3);
7

```

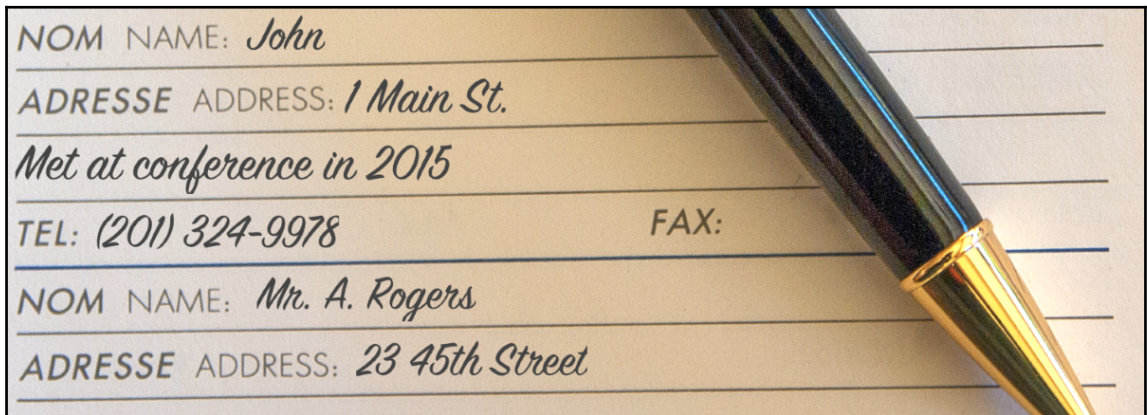
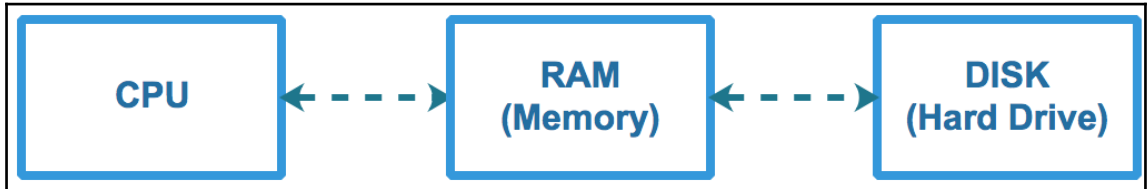
Results (28)

	y.alpha3	y.latitude	y.longitude	x.location	x.avgprice
1	AUS	-27	133	AUS	43.655062157137991
2	AUT	47.333301544189453	13.33329963684082	AUT	25.300277796658602
3	BEL	50.833301544189453	4	BEL	40.222350115907446
4	CAN	60	-95	CAN	41.546758076452441
5	CHE	47	8	CHE	37.297826310862668
6	CZE	49.75	15.5	CZE	75.050954674955591



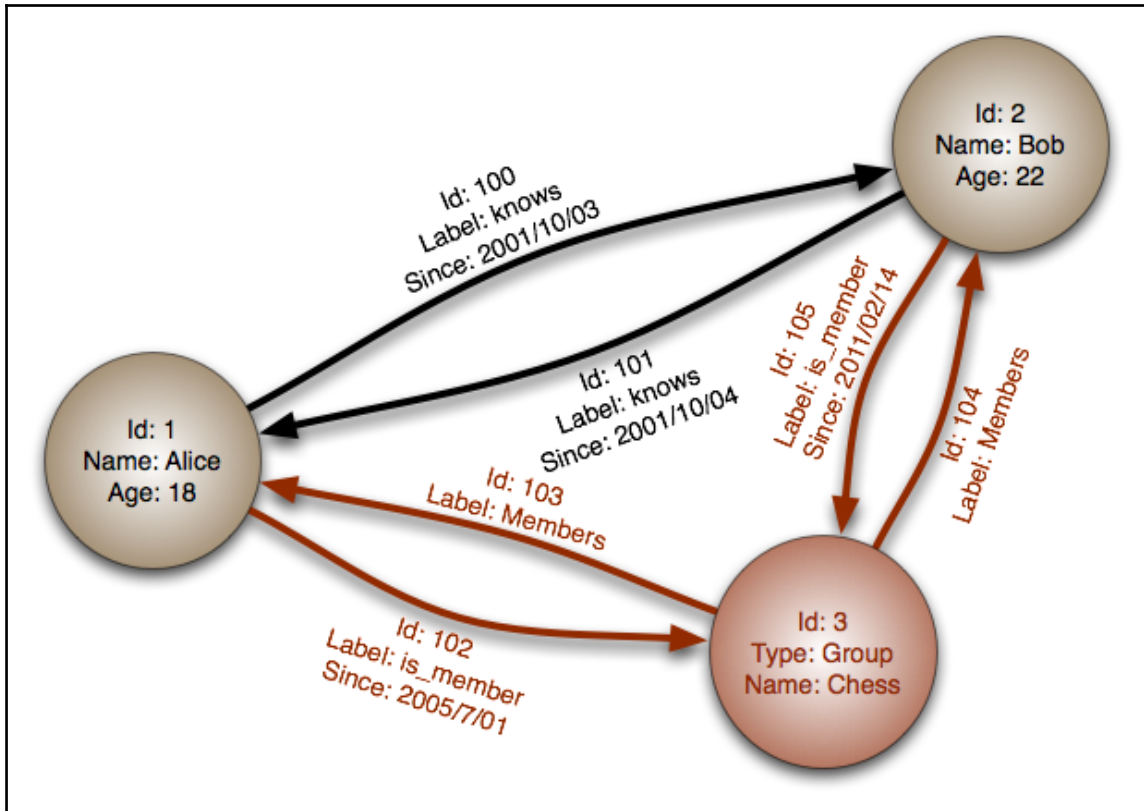


Chapter 5: Big Data Mining with NoSQL



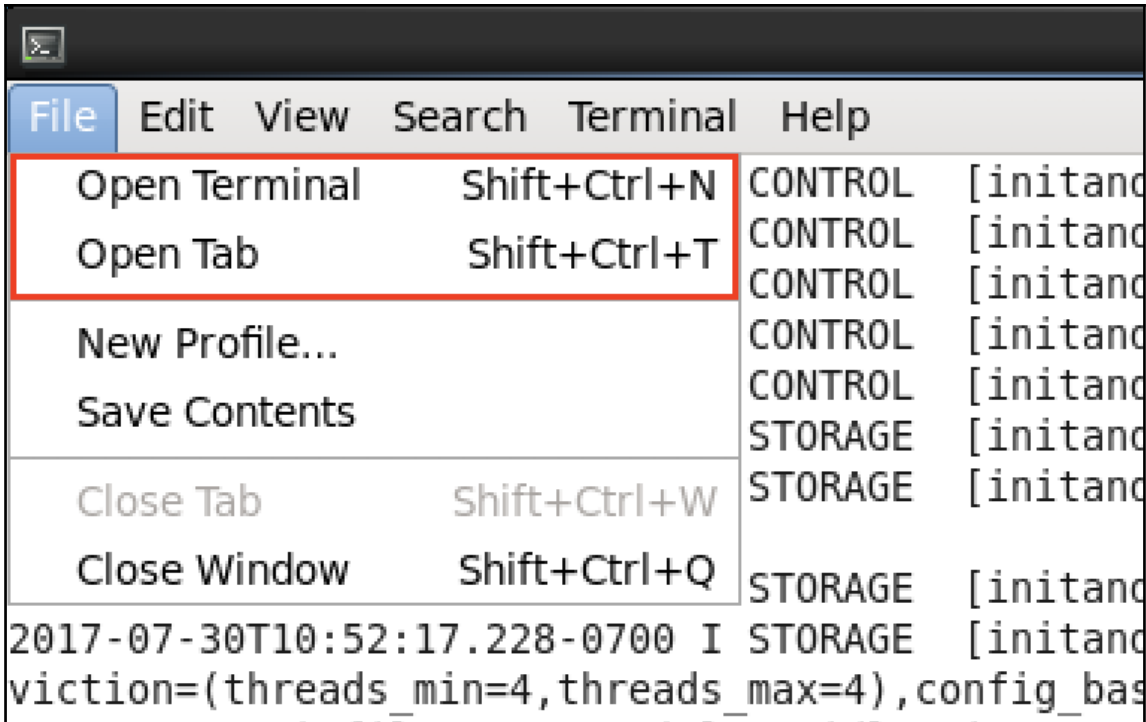
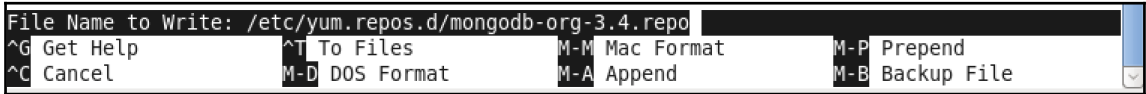
A	: 1	B	: 2	C	: 3	D	: 4	E	: 5	F	: 6
G	: 7	H	: 8	I	: 9	J	: 10	K	: 11	L	: 12
M	: 13	N	: 14	O	: 15	P	: 16	Q	: 17	R	: 18
S	: 19	T	: 20	U	: 21	V	: 22	W	: 23	X	: 24
Y	: 25	Z	: 26								

H		8
A		1
M		13
L		12
E		5
T		20



```

root@quickstart:/home/cloudera
File Edit View Search Terminal Help
GNU nano 2.0.9 File: /etc/yum.repos.d/mongodb-org-3.4.repo Modified
[mongodb-org-3.4]
name=MongoDB Repository
baseurl=https://repo.mongodb.org/yum/redhat/$releasever/mongodb-org/3.4/x86_64/
gpgcheck=1
enabled=1
gpgkey=https://www.mongodb.org/static/pgp/server-3.4.asc
^G Get Help      ^O WriteOut     ^R Read File    ^Y Prev Page    ^K Cut Text      ^C Cur Pos
^X Exit          ^J Justify      ^W Where Is     ^V Next Page    ^U UnCut Text   ^T To Spell
    
```



```
File Edit View Search Terminal Help
GNU nano 2.0.9 File: laureates.json
{"laureates":[{"id":"1","firstname":"Wilhelm Conrad","su
```

We will delete {"laureates": from the beginning of the file and the letter } from the end of the file

```
    "gender" : "male",
    "prizes" : [
      {
        "year" : "1901",
        "category" : "physics",
        "share" : "1",
        "motivation" : "\"in recognition of his services he has rendered by the discovery of
        presently named after him\"",
        "affiliations" : [
```

category is a field nested in the field prizes

DOCS UNIVERSITY WHAT'S MONGODB? LOGIN

[Free Sandbox](#)
[Download](#)

mongoDB | FOR GIANT IDEAS
SOLUTIONS CLOUD CUSTOMERS RESOURCES ABOUT US

Try MongoDB Atlas,
our new database as a service

[Start Free](#)

Enterprise Server
Ops Manager
Compass
Connector for BI

As the GUI for MongoDB, MongoDB Compass allows you to make smarter decisions about indexing, document validation, and more. Compass enables you to:

- Visually explore the structure of data in your database
- Run ad hoc queries in seconds
- View and optimize your query performance
- Interact with your data with full CRUD functionality
- Create and modify rules that validate your data
- Compass is available as part of the MongoDB Professional and MongoDB Enterprise Advanced subscriptions

Versions:

Platforms:

[Download](#)

[Documentation](#)

NEW CONNECTIONS
FAVORITES RECENT CONNECTIONS

Welcome to MongoDB Compass

The charts show database operations, queues, network traffic, connections and memory over time.

The percentage value represents the system load for each collection.

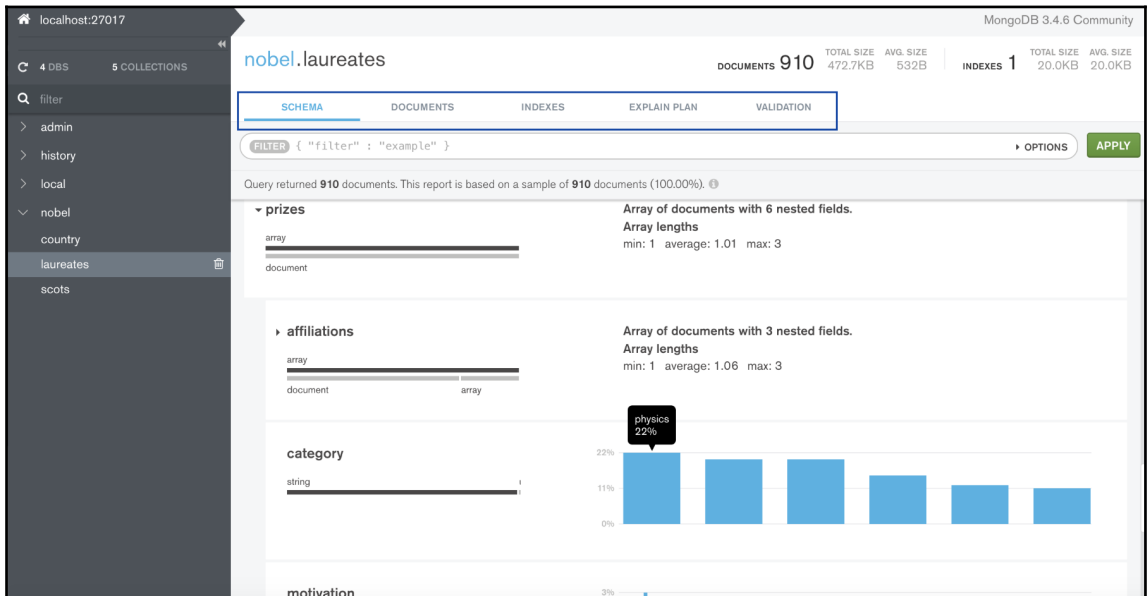
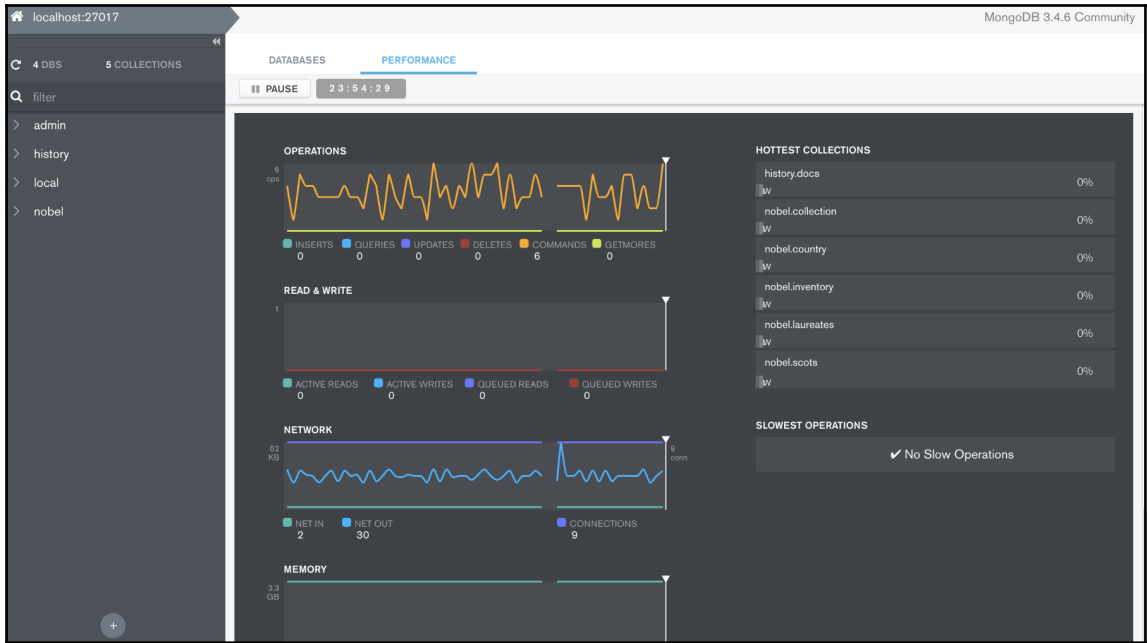
Collection	System Load
ms-cb-1 coll2	80.01%
ms-cb-3 coll1	80.90%
ms-cb-1 coll3	1.81%
test.db.coll	0%
database.zipcodes	0%
ms-cb-1 coll1	0%

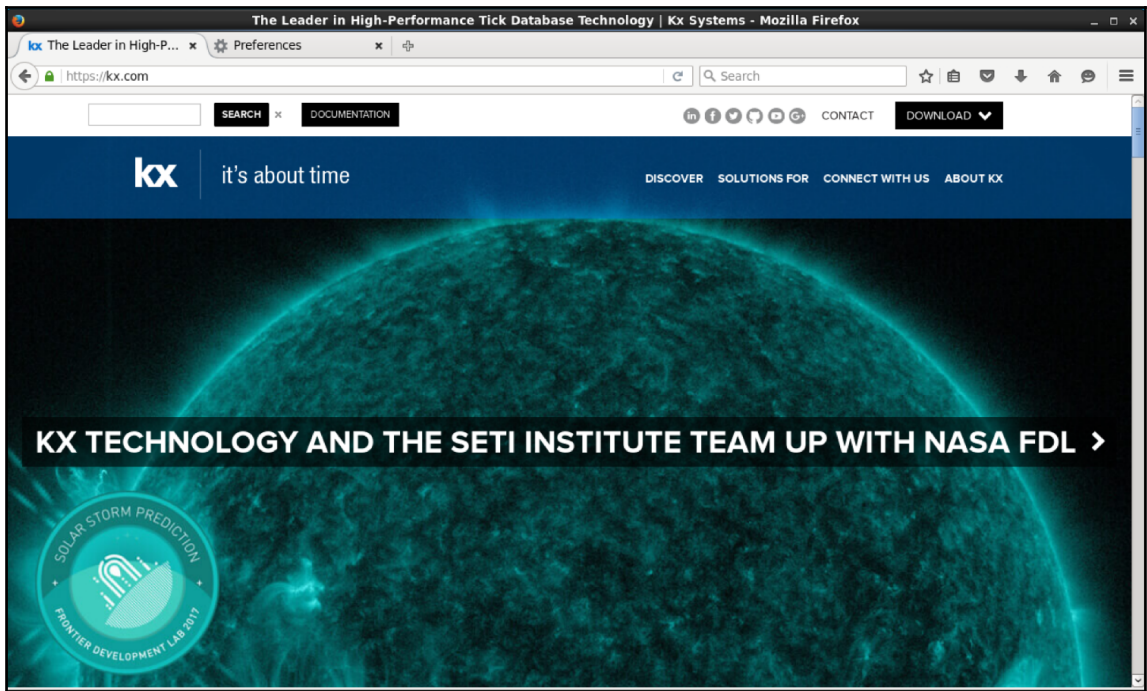
Click on a slow operation to examine more details.

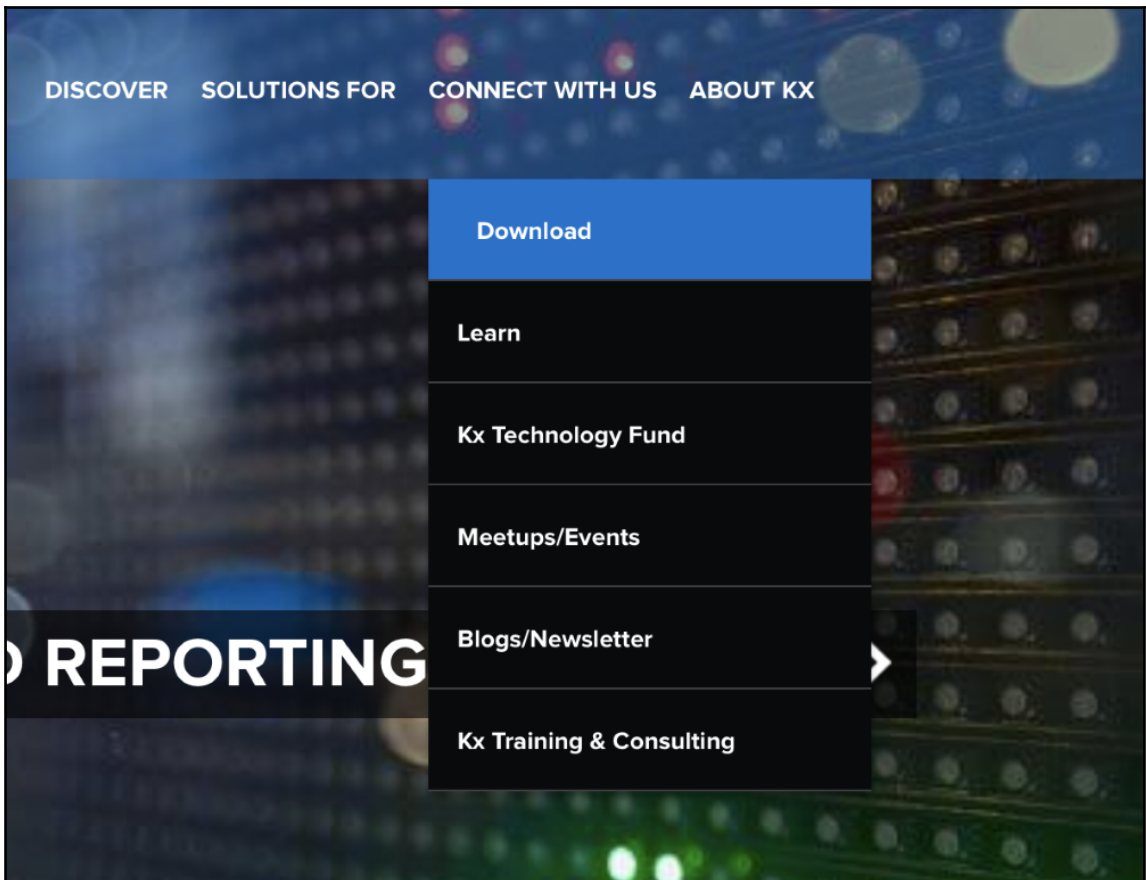
Performance Charts.

Real-time server statistics let you view key server metrics and database operations. Drill down into database operations easily and understand your most active collections.

Next >







The screenshot shows the top of a web browser with a dark blue header. On the left is the 'kx' logo, and on the right is the tagline 'it's about time'. Below the header, the main content area has a dark, textured background. At the top center, the text 'DOWNLOAD KDB+' is displayed in large, white, sans-serif font. Below this, a white rectangular box contains the following text: 'No Fee, Non-Commercial Use, 32 Bit Kdb+ Software License Agreement (rev. 13 Sep 2015) CAREFULLY READ THE FOLLOWING TERMS AND CONDITIONS. BY DOWNLOADING THE 32 Bit KDB+ SOFTWARE, YOU ARE AGREEING TO BE BOUND BY THESE TERMS AND CONDITIONS. IF YOU DO NOT AGREE TO THESE TERMS AND CONDITIONS, DO NOT DOWNLOAD THE SOFTWARE. This No Fee, Non-Commercial Use, 32 Bit Kdb+ Software License Agreement ("Agreement") is made between Kx'. Below the license agreement is a white input field for an email address, with the placeholder text '(Enter your email address here)'. Underneath the email field are two checkboxes: the first is checked and labeled 'I agree to the terms of the license agreement for the free version of kdb+ (which is 32-bit vs. 64-bit for the commercial version)', and the second is unchecked and labeled 'I would like to receive the Quarterly Kx Newsletter'. Below the checkboxes are four rounded rectangular buttons: 'Windows' (yellow-green), 'Linux-x86' (green), 'Linux-ARM' (purple), and 'OSX' (blue). At the bottom right of the main content area, there is a link that says 'No, thanks. Take me to the [Kx Systems Homepage](#)'.

The screenshot shows the bottom of a web browser window. The address bar contains a search icon, a search input field with the text 'Search', and icons for star, list, shield, and download. Below the address bar, the download bar is visible, showing a file named 'linuxx86.zip' with a folder icon to its left. Below the filename, the file size '312 KB' and the source 'kx.com' are displayed, followed by the time '09:28 AM'. A small trash can icon is visible on the right side of the download bar.

The Comprehensive R Archive Network

Download and Install R

Precompiled binary distributions of the base system and contributed packages, **Windows and Mac** users most likely want one of these versions of R:

- [Download R for Linux](#)
- [Download R for \(Mac\) OS X](#)
- [Download R for Windows](#)

R is part of many Linux distributions, you should check with your Linux package management system in addition to the link above.

RStudio - RStudio - Mozilla Firefox

Free downloa... x RStudio - RStudio x

https://www.rstudio.com/products/rstudio/

RStudio®

rstudio::conf Products Resources Pricing About Us Blogs

RStudio Desktop

Open Source Edition Commercial License

Overview

- Access RStudio locally
- Syntax highlighting, code completion, and smart indentation
- Execute R code directly from the source editor
- Quickly jump to function definitions
- Easily manage multiple working directories using projects
- Integrated R help and documentation
- Interactive debugger to diagnose and fix errors quickly
- Extensive package development tools

All of the features of open source; plus:

- A commercial license for organizations not able to use AGPL software
- Access to priority support

R Studio [Products](#) [Resources](#) [Pricing](#) [About Us](#) [Blogs](#) [rstudio::conf](#)

- Integrated R help
- Interactive debugger to diagnose and fix errors quickly
- Extensive package development tools

Support Community forums only

- Priority Email Support
- 8 hour response during business hours (ET)

License AGPL v3 [RStudio License Agreement](#)

Pricing Free \$995/year

[DOWNLOAD RSTUDIO DESKTOP](#) [BUY NOW](#)


R Studio [Products](#) [Resources](#) [Pricing](#) [About Us](#) [Blogs](#) [rstudio::conf](#)

Choose Your Version of RStudio

RStudio is a set of integrated tools designed to help you be more productive with R. It includes a console, syntax-highlighting editor that supports direct code execution, and a variety of robust tools for plotting, viewing history, debugging and managing your workspace. [Learn More about RStudio features.](#)

License Type	Price
RStudio Desktop Open Source License	FREE
RStudio Desktop Commercial License	\$995 per year
RStudio Server Open Source License	FREE
RStudio Server Pro Commercial License	\$9,995 per year
RStudio Server Pro + RStudio Connect Commercial License	\$29,995 per year

Integrated Tools for R



[rstudio::conf](#)
[Products](#)
[Resources](#)
[Pricing](#)
[About Us](#)

[RStudio Desktop 1.0.153 -- Release Notes](#)

RStudio requires R 2.11.1+. If you don't already have R, download it [here](#).

Installers for Supported Platforms

Installers	Size	Date	MD5
RStudio 1.0.153 - Windows Vista/7/8/10	81.9 MB	2017-07-20	b3b4bbc82865ab105c21cb70b17271b3
RStudio 1.0.153 - Mac OS X 10.6+ (64-bit)	71.2 MB	2017-07-20	8773610566b74ec3e1a88b2fdb10c8b5
RStudio 1.0.153 - Ubuntu 12.04-15.10/Debian 8 (32-bit)	85.5 MB	2017-07-20	981be44f91fc07e5f69f52330da32659
RStudio 1.0.153 - Ubuntu 12.04-15.10/Debian 8 (64-bit)	91.7 MB	2017-07-20	2d0769bea2bf6041511d6901a1cf69c3
RStudio 1.0.153 - Ubuntu 16.04+/Debian 9+ (64-bit)	61.9 MB	2017-07-20	d584cbab01041777a15d62cbe69a976
RStudio 1.0.153 - Fedora 19+/RedHat 7+/openSUSE 13.1+ (32-bit)	84.7 MB	2017-07-20	8dfee96059b05a063c49b705eca0ceb4
RStudio 1.0.153 - Fedora 19+/RedHat 7+/openSUSE 13.1+ (64-bit)	85.7 MB	2017-07-20	16c2c8334f961c65d9bfa8fb813ad7e7

Zip/Tarballs

Zip/tar archives	Size	Date	MD5
RStudio 1.0.153 - Windows Vista/7/8/10	117.6 MB	2017-07-20	024b5714fa6ef337fe0c6f5e2894cbcb
RStudio 1.0.153 - Ubuntu 12.04-15.10/Debian 8 (32-bit)	86.2 MB	2017-07-20	f8e0ffa7ec62665524f9e2477facd346
RStudio 1.0.153 - Ubuntu 12.04-15.10/Debian 8 (64-bit)	92.7 MB	2017-07-20	2077c181311d1aad6fb8d435f8f1f45f
RStudio 1.0.153 - Fedora 19+/RedHat 7+/openSUSE 13.1+ (32-bit)	85.4 MB	2017-07-20	92e1a22d14952273ec389e5a55be614f
RStudio 1.0.153 - Fedora 19+/RedHat 7+/openSUSE 13.1+ (64-bit)	86.6 MB	2017-07-20	0b71c5a7fc53c84b3fe67242240b3531

The screenshot shows a web browser window with the URL `openpaymentsdata.cms.gov`. The page title is "OpenPaymentsData.CMS.gov". In the top right corner, there are links for "About" and "Sign In". The main content area features a large banner with the heading "Find Your Doctor's Payments" and a search form with fields for "First Name" and "Last Name", a "Search" button, and a yellow button labeled "Explore Open Payments Data". Below the search form, there is a paragraph of text: "Looking for a teaching hospital or company making payments? For a more refined search, use the Search Tool." followed by two links: "How to use the Search Tool" and "Check out the new Payments by State page".

Below the banner is a section titled "The Facts on Open Payments". It contains two data points:

2013 Totals	2014 Totals	2015 Totals	2016 Totals	Totals All Years
			8.18 BILLION	
				11.96 MILLION

The first row of data is labeled "Total US Dollar Value" and the second row is labeled "Total Records Published". There are also two buttons: "See Summary Data" and "See Payments by State page".

OpenPaymentsData.CMS.gov

🏠 | About | Sign In

Search Tool | Summary Data | Data Explorer | Download Data | Open Payments Home

Sign In to OpenPaymentsData.CMS.gov

✉ Email

🔒 Password

[Forgot Password?](#)

Sign In

Don't have an account yet? [Sign Up](#)

Search ToolSummary DataData ExplorerDownload DataOpen Payments Home

We're glad you want to join OpenPaymentsData.CMS.gov

Create a new Socrata ID.

Use your Email and Password to sign in to all Socrata powered sites.

Email Address * Please enter a valid email address.


Display Name *

Password * Restrictions apply

Confirm Password *

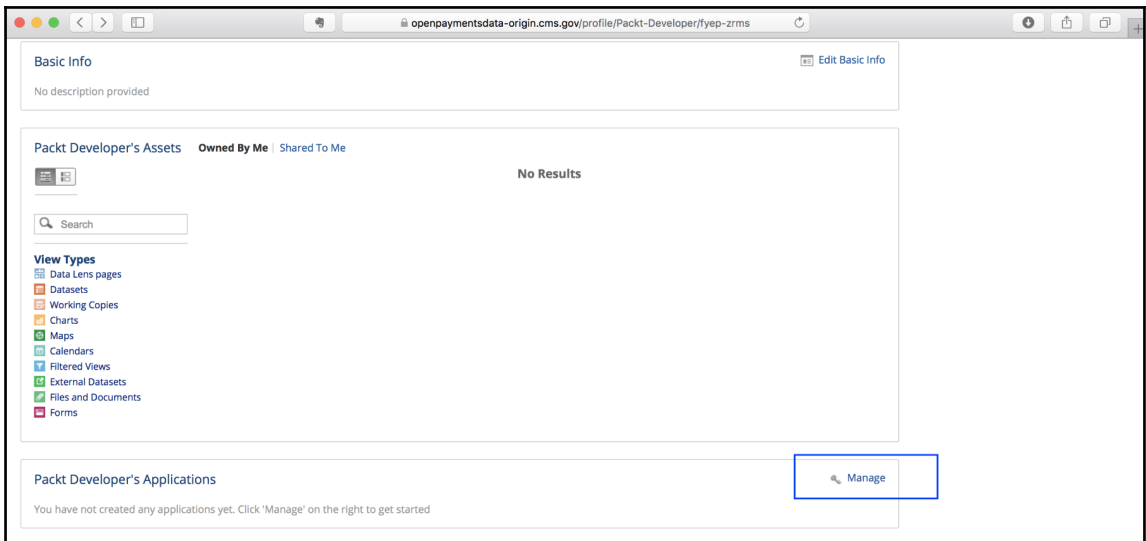
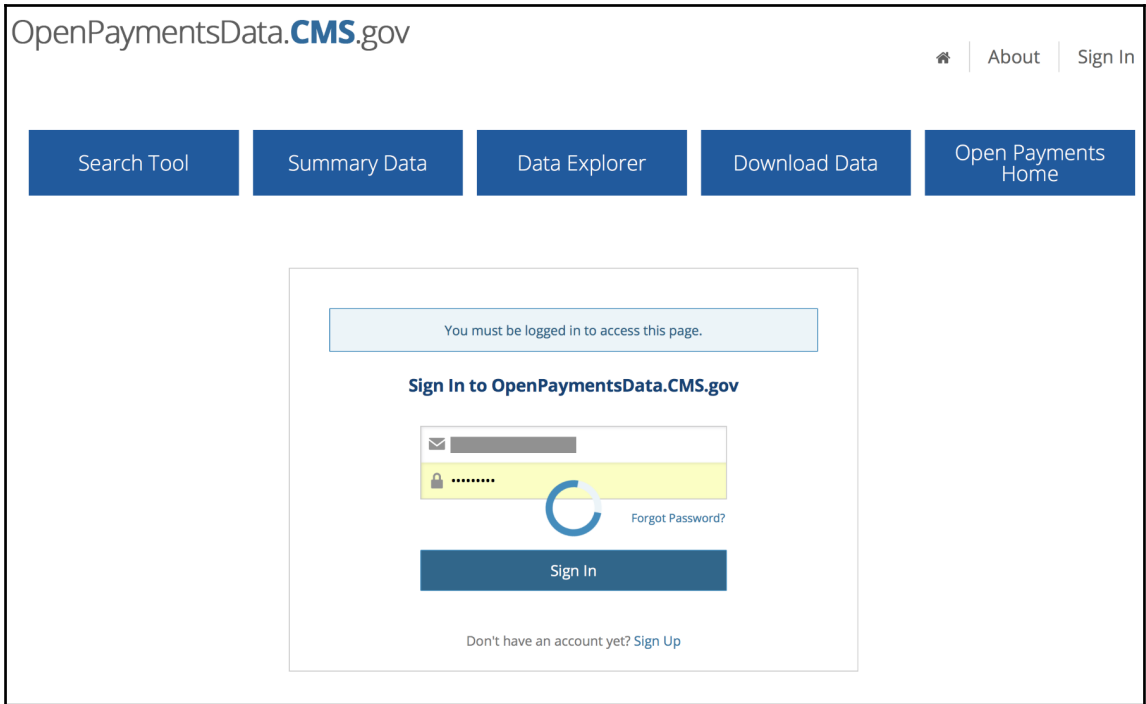
✓

I'm not a robot


reCAPTCHA
Privacy - Terms

By clicking on "Create My Account" below, you are agreeing to our [Terms of Service](#) and [Privacy Policy](#).

Create My Account



OpenPaymentsData.CMS.gov

[About](#)
[Profile](#)

Search Tool

Summary Data

Data Explorer

Download Data

Open Payments Home

- Basic Info
- Profile Image
- Account Settings
- App Tokens

Your Applications

- * **Application Name**
- * **Description**
- Organization**
- Website**
- Callback Prefix**

For instance, if your application will call back to https://my-domain.com/socrata-app/oauth/validated and https://my-domain.com/socrata-app/profile, you'll want to supply https://my-domain.com/socrata-app as your prefix. Note that you must include at least the entire domain name.
- Public?**

Create
Cancel

Edit Application Token

Your application has been created

- * **Application Name**
- * **Description**
- Organization**
- Website**
- Callback Prefix**

Basic Info

Profile Image

Account Settings

App Tokens

Your Applications

Create New Application

<i>No Thumbnail</i>	Name	CMS Open Payments 2016 Web Portal
	Description	A Web Portal for CMS Open Payments Data
	App Token	dciikbI3jF1Ds6weHZmQmsQt0
	Secret Token	Show Secret Token

Edit **Delete**

```
virtual_machine_default_1503274187494_79006 [Running]
Ubuntu 16.04.2 LTS vagrant tty1
vagrant login:
Ubuntu 16.04.2 LTS vagrant tty1
vagrant login: packt
Password:
Welcome to Ubuntu 16.04.2 LTS (GNU/Linux 4.4.0-83-generic x86_64)

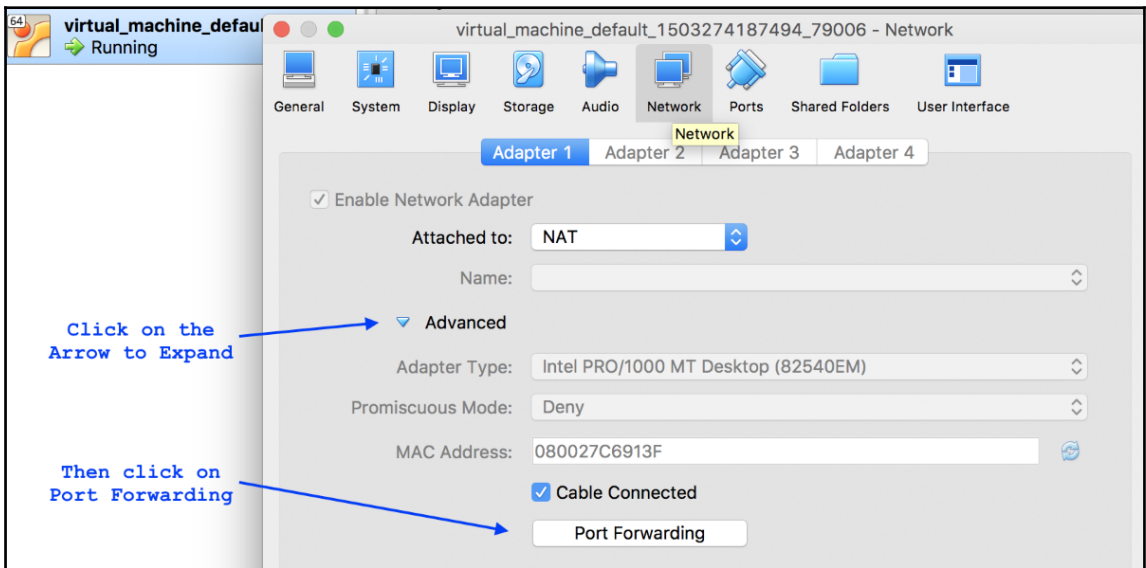
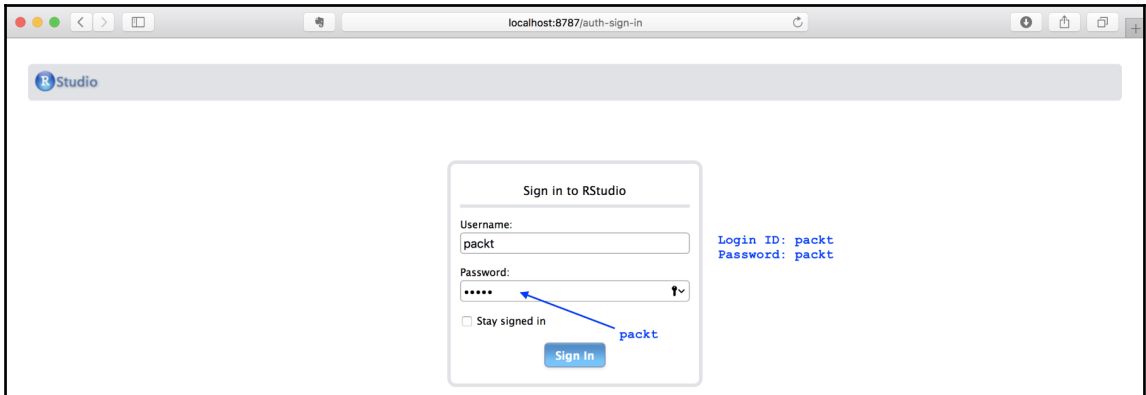
 * Documentation:  https://help.ubuntu.com
 * Management:    https://landscape.canonical.com
 * Support:       https://ubuntu.com/advantage

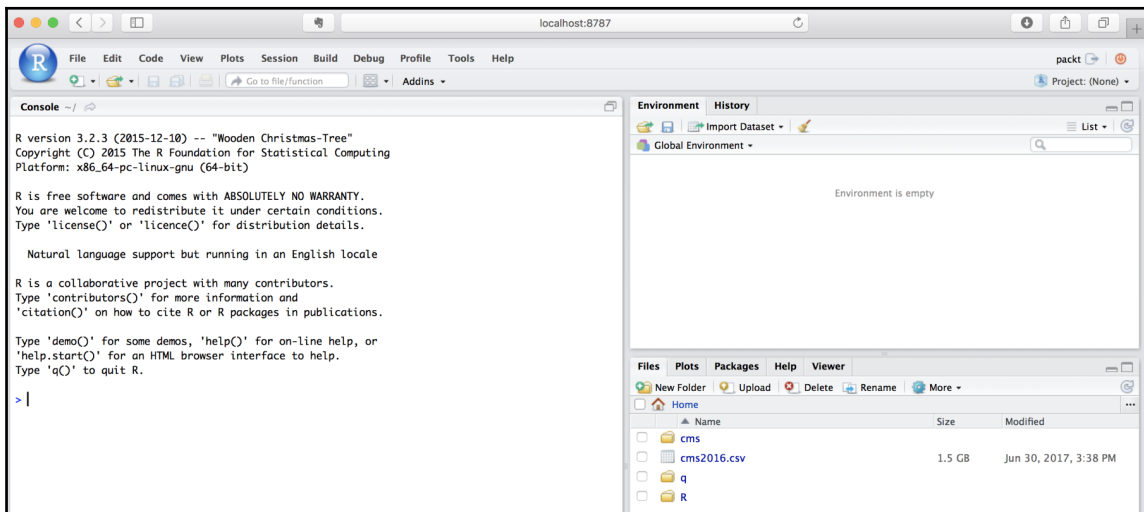
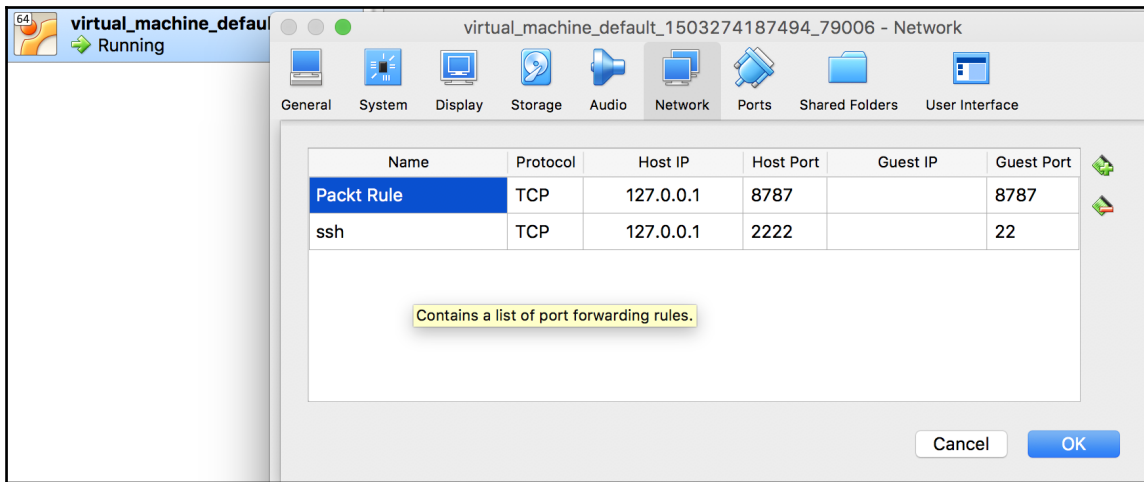
38 packages can be updated.
6 updates are security updates.

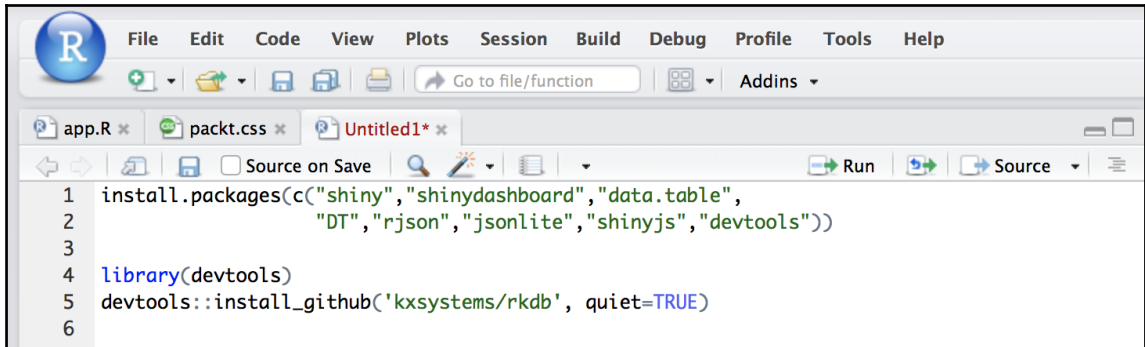
The programs included with the Ubuntu system are free software;
the exact distribution terms for each program are described in the
individual files in /usr/share/doc/*/copyright.

Ubuntu comes with ABSOLUTELY NO WARRANTY, to the extent permitted by
applicable law.

packt@vagrant:~$ _
```

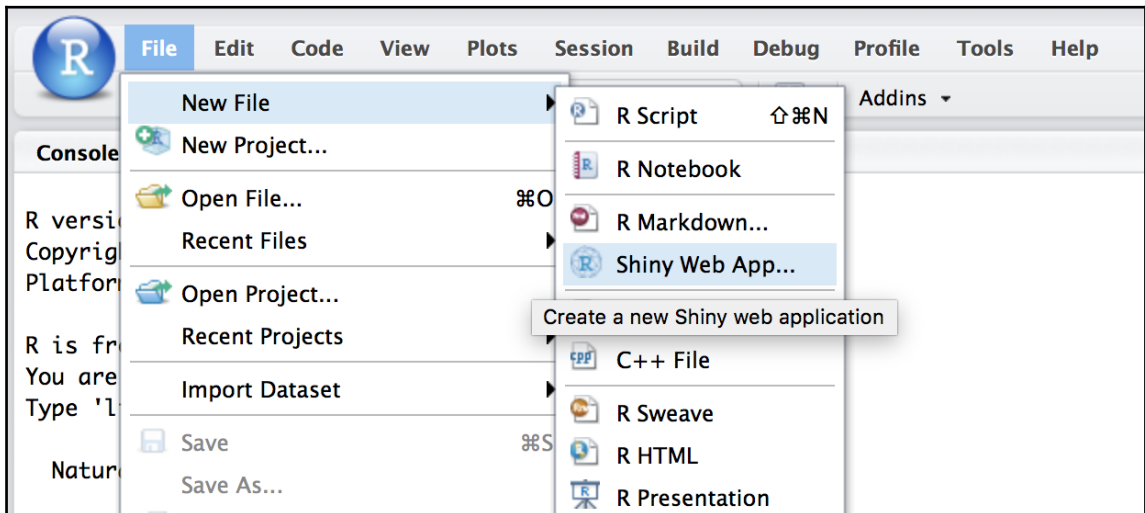


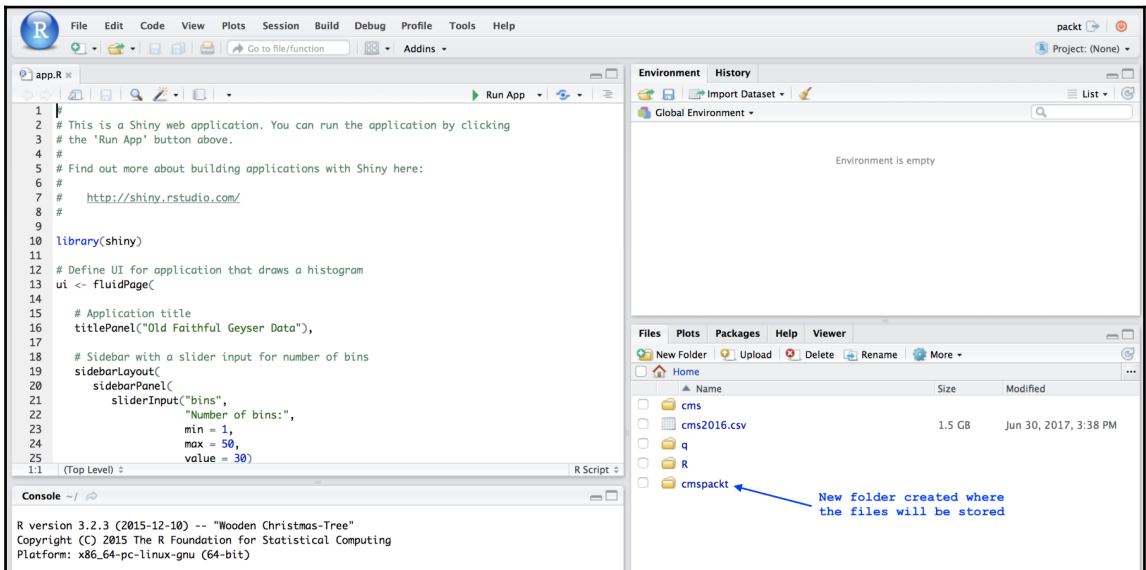
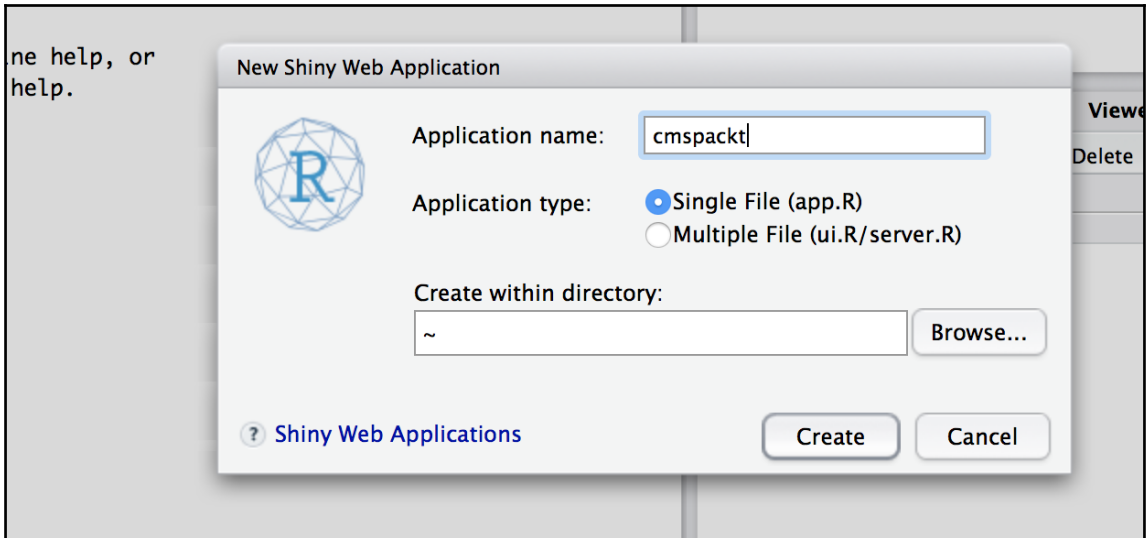


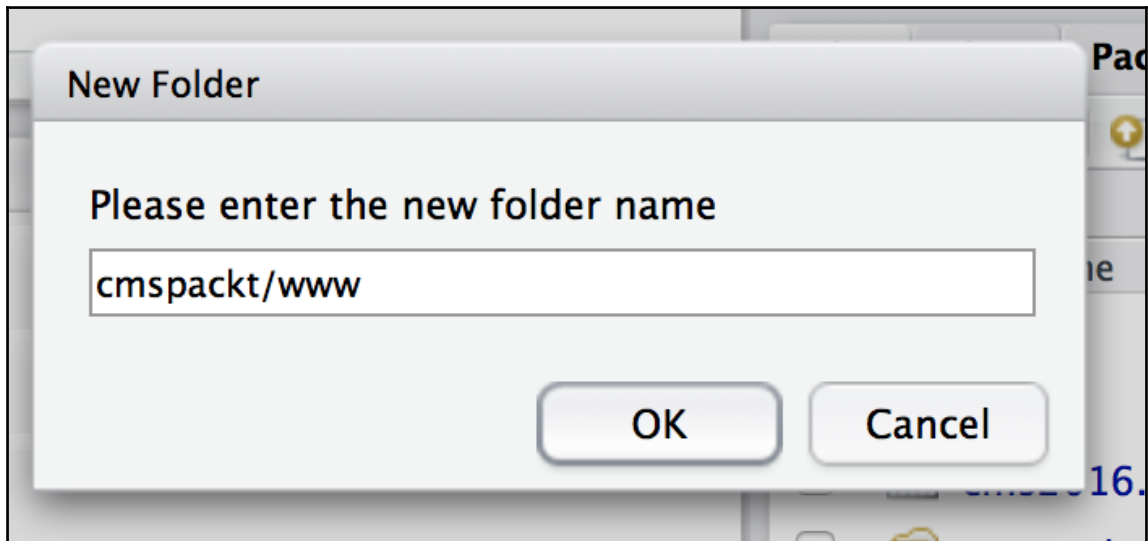
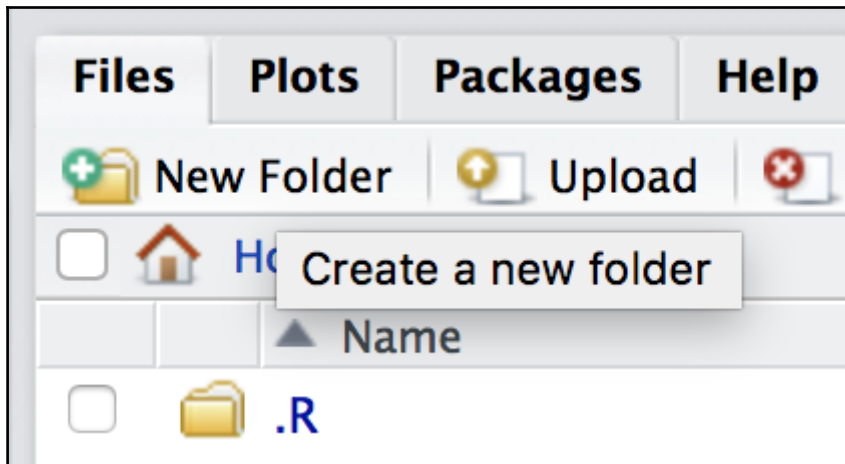


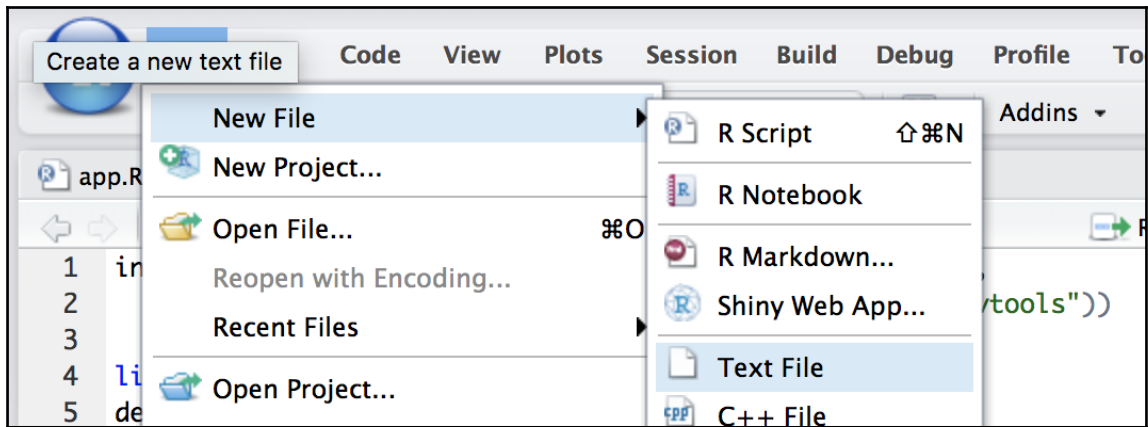
The screenshot shows the RStudio editor window with the following R code:

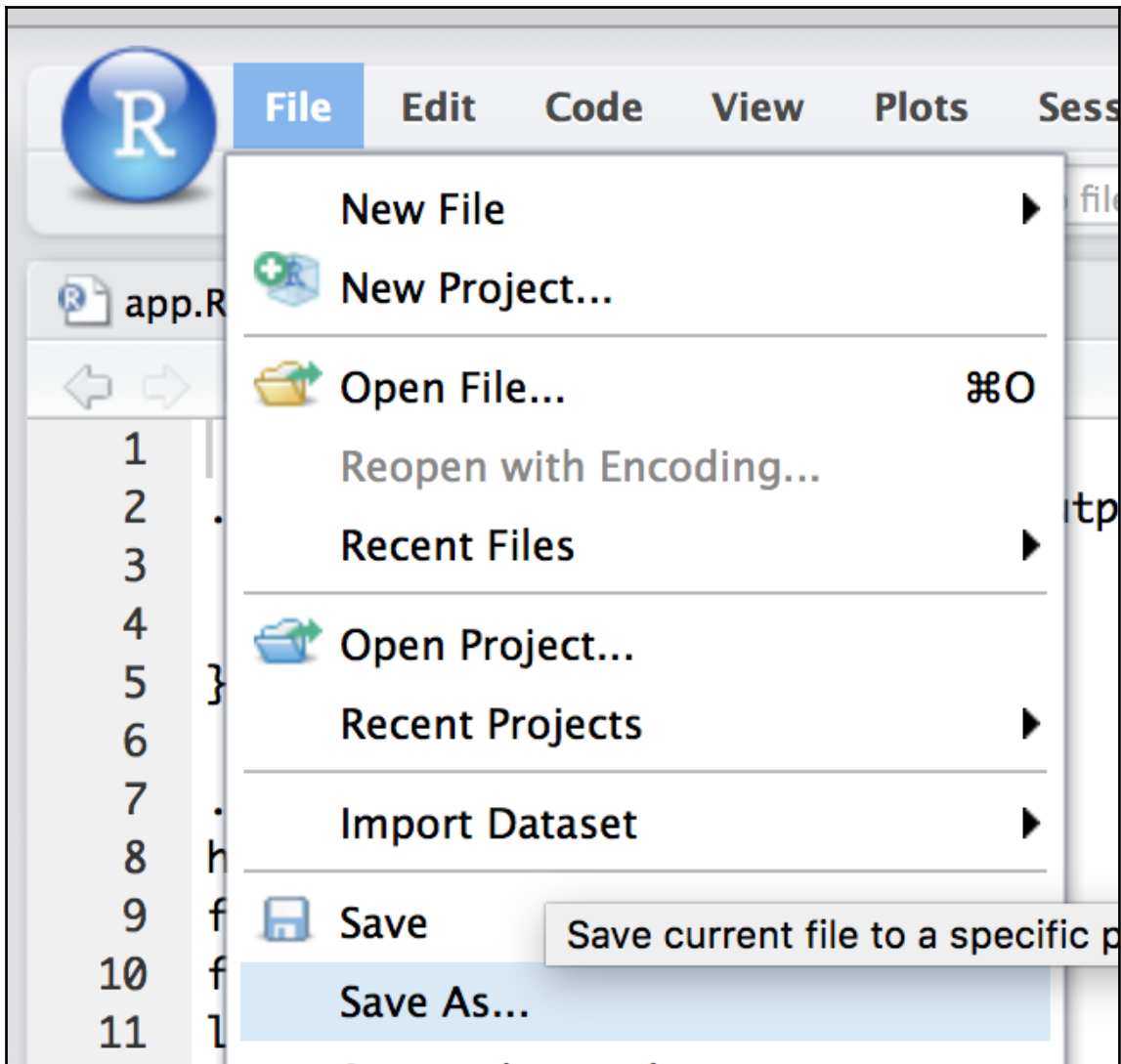
```
1 install.packages(c("shiny", "shinydashboard", "data.table",  
2                   "DT", "rjson", "jsonlite", "shinyjs", "devtools"))  
3  
4 library(devtools)  
5 devtools::install_github('kxsystems/rkdb', quiet=TRUE)  
6
```

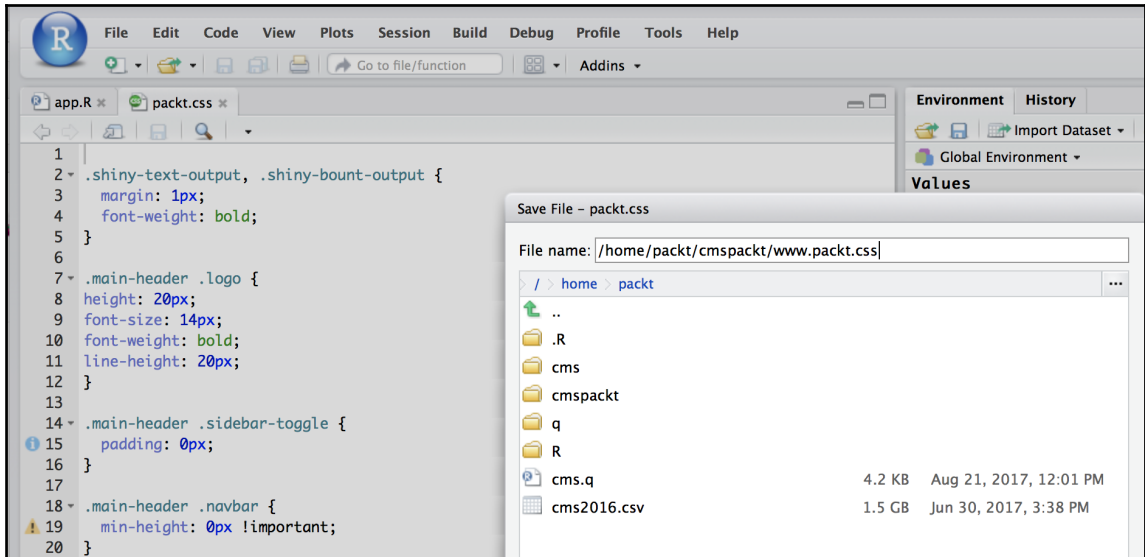










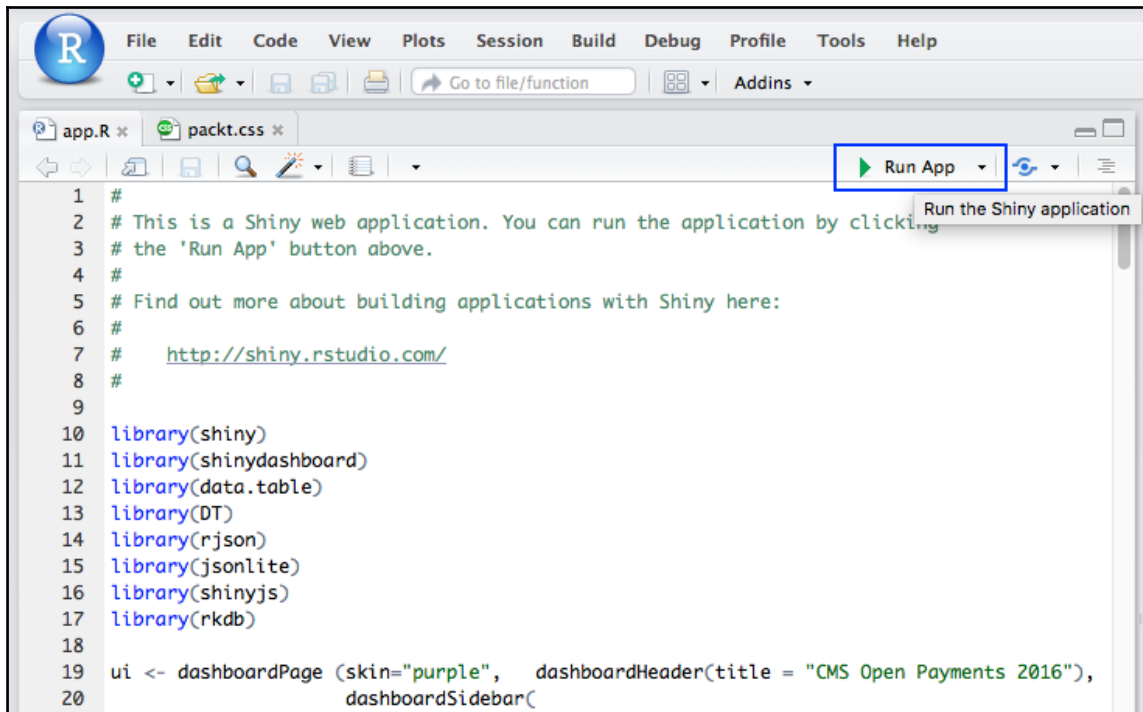


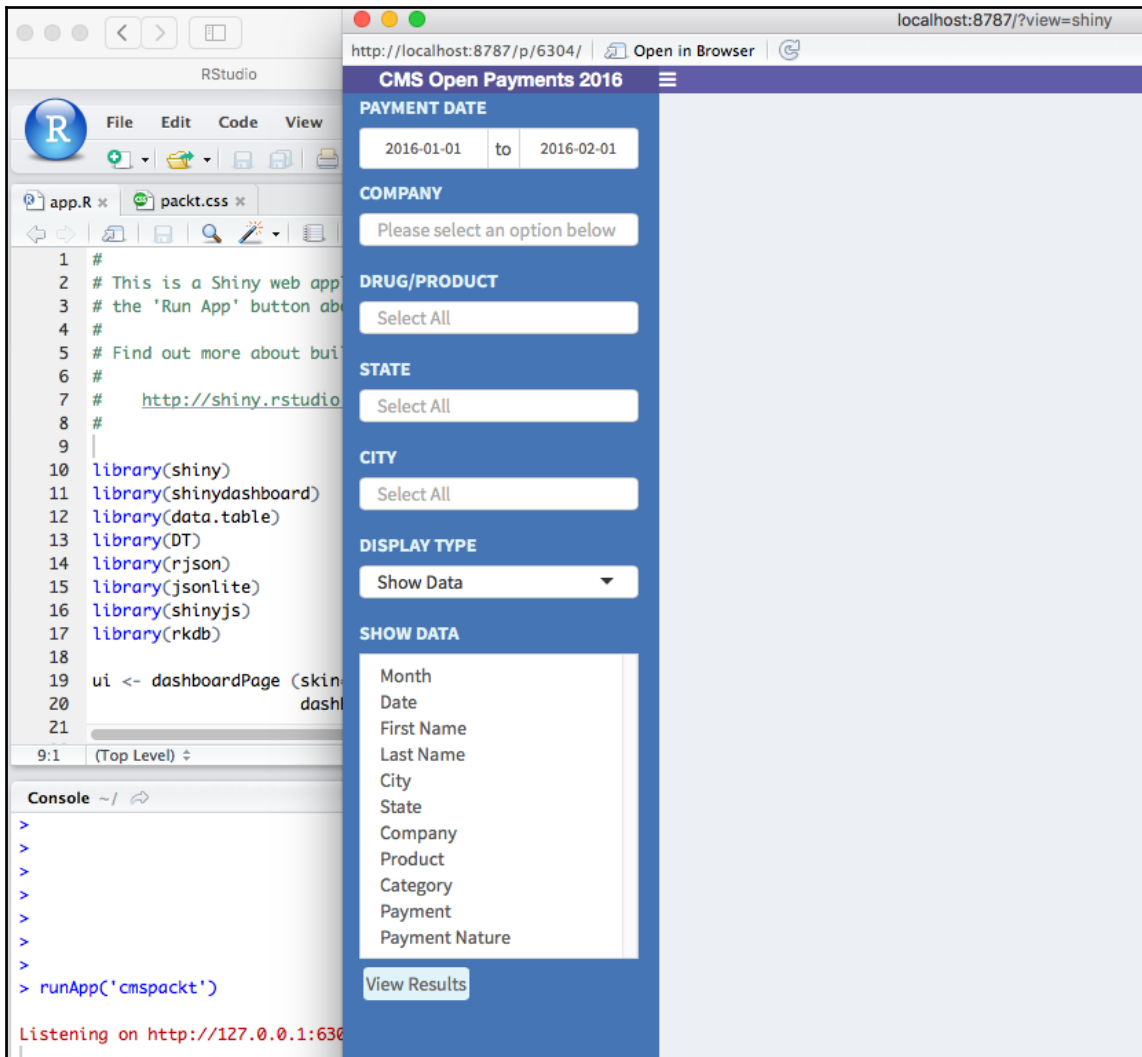
```

packt@vagrant:~$
packt@vagrant:~$ pwd
/home/packt
packt@vagrant:~$ q
KDB+ 3.5 2017.06.15 Copyright (C) 1993-2017 Kx Systems
132/ 1()core 3951MB packt vagrant 127.0.1.1 NONEXPIRE

Welcome to kdb+ 32bit edition
For support please see http://groups.google.com/d/forum/personal-kdbplus
Tutorials can be found at http://code.kx.com/wiki/Tutorials
To exit, type \
To remove this startup msg, edit q.q
q)\1 cms.q
q)

```





CMS Open Payments 2016

16:58:17.273: Processed 11,258,311 records in 00:00:00.021 seconds. Returned result with 172 rows.

Show 10 entries

Search:

	Month	Date	Payment
1	2016.01	2016-01-01	13.5
2	2016.01	2016-01-04	3.62
3	2016.01	2016-01-04	7.77
4	2016.01	2016-01-04	4220
5	2016.01	2016-01-04	4.73
6	2016.01	2016-01-04	14.05
7	2016.01	2016-01-04	18.9
8	2016.01	2016-01-04	5170
9	2016.01	2016-01-04	2.66
10	2016.01	2016-01-04	11.63

Showing 1 to 10 of 172 entries

Previous 2 3 4 5 ... 18 Next

SHOW DATA

- Month
- Date
- First Name
- Last Name
- City
- State
- Company
- Product
- Category
- Payment
- Payment Nature

[View Results](#)

16:52:46.354: Processed 11,258,311 records in 00:00:00.022 seconds. Returned result with 116 rows.

16:52:46.354: Processed 11,258,311 records in 00:00:00.022 seconds. Returned result with 116 rows.

Show 10 entries

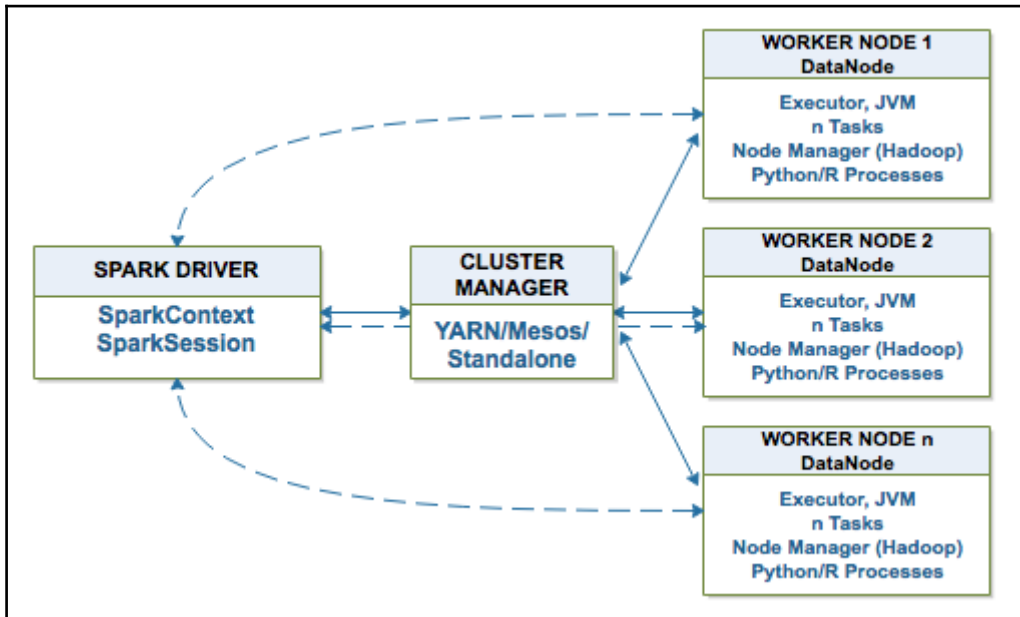
Search:

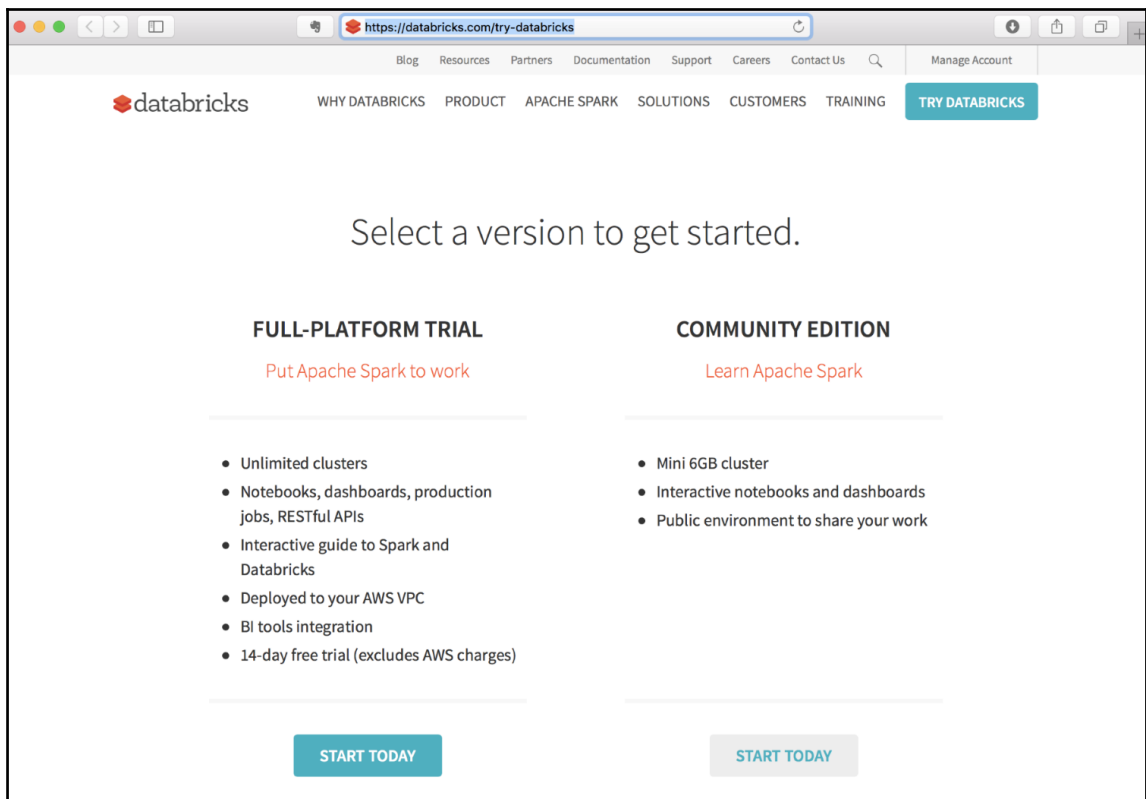
	State	Category	PaymentNature	Total Payment
1	ak	gastrointestinal;inflammation;ophthalmology;transplant	food and beverage	72.33
2	al	gastrointestinal;inflammation;ophthalmology;transplant	consulting fee	4687.5
3	al	gastrointestinal;inflammation;ophthalmology;transplant	food and beverage	182.57
4	al	gastrointestinal;inflammation;ophthalmology;transplant	travel and lodging	550.17
5	ar	gastrointestinal;inflammation;ophthalmology;transplant	food and beverage	294.59
6	az	gastrointestinal;inflammation;ophthalmology;transplant	compensation for services other than consulting, including serving as faculty or as a speaker at a venue other than a continuing education program	3600
7	az	gastrointestinal;inflammation;ophthalmology;transplant	food and beverage	293.52
8	ca	gastrointestinal;inflammation;ophthalmology;transplant	compensation for services other than consulting, including serving as faculty or as a speaker at a venue other than a continuing education program	12180
9	ca	gastrointestinal;inflammation;ophthalmology;transplant	consulting fee	29400
10	ca	gastrointestinal;inflammation;ophthalmology;transplant	food and beverage	3861.95

Showing 1 to 10 of 116 entries

Previous ... Next

Chapter 6: Spark for Big Data Analytics





The screenshot shows a web browser window with the URL <https://databricks.com/try-databricks>. The page features the Databricks logo and a navigation menu with links for Blog, Resources, Partners, Documentation, Support, Careers, Contact Us, and Manage Account. Below the navigation, there are links for WHY DATABRICKS, PRODUCT, APACHE SPARK, SOLUTIONS, CUSTOMERS, TRAINING, and a prominent blue button labeled TRY DATABRICKS.

The main content area is titled "Select a version to get started." and is divided into two columns:

- FULL-PLATFORM TRIAL**
Put Apache Spark to work
- COMMUNITY EDITION**
Learn Apache Spark

Each column contains a list of features and a "START TODAY" button.

FULL-PLATFORM TRIAL Features:

- Unlimited clusters
- Notebooks, dashboards, production jobs, RESTful APIs
- Interactive guide to Spark and Databricks
- Deployed to your AWS VPC
- BI tools integration
- 14-day free trial (excludes AWS charges)

COMMUNITY EDITION Features:

- Mini 6GB cluster
- Interactive notebooks and dashboards
- Public environment to share your work

The image shows a web browser window with the URL `accounts.cloud.databricks.com/registration.html#signup/community`. The page title is "Sign Up for Databricks Community Edition". The form contains the following fields:

- First Name ***: Input field with "John" and a yellow placeholder "ENTER FIRST NAME".
- Last Name ***: Input field with "Doe" and a yellow placeholder "ENTER LAST NAME".
- Company Name ***: Input field with "NA".
- Work Email ***: Input field with "email@email.com" and a yellow placeholder "ENTER YOUR EMAIL ADDRESS HERE".
- Password ***: Input field with "*****" and a yellow placeholder "ENTER A PASSWORD".
- Confirm Password ***: Input field with "*****" and a yellow placeholder "CONFIRM YOUR PASSWORD".
- Phone Number**: Empty input field.
- What is your intended use case? ***: Input field with "Personal - Learning Spark".
- How would you describe your role? ***: Input field with "Data Scientist".

At the bottom left, there is a reCAPTCHA widget with a green checkmark and the text "I'm not a robot". At the bottom right, there is a teal "Sign Up" button.

Terms of Service

Read and scroll to the bottom to continue ↓




Databricks Community Edition

Terms of Service


(Posted May 24, 2016)

Welcome to Databricks Community Edition! We are pleased to provide Databricks Community Edition (**“Community Edition”**) at no charge to those interested in learning and exploring the use of Databricks’ cloud-based data analytics platform, which enables data analysts and others to easily tap the power of Apache Spark™ and Databricks’ other proprietary functionality. Your use of Community Edition, whether on an ongoing basis or on a temporary basis for the purposes of trialing Databricks’ fee-based platform services, or your temporary, no-charge trial of our fee-based platform services apart from Community Edition (**“Free Trial”**) and, together with any use of Community Edition, the **“Services”**) is governed by these Terms of Service (the **“Terms”**). By using the Services, you are agreeing to be bound by these Terms. If you are using the Services on behalf of an organization, you are agreeing to these Terms on behalf of that organization and you are confirming that you are authorized to bind it to these Terms (in which case **“you”** or **“your”** refer to that organization rather than you as an individual). Your access to the Services is contingent upon your reading these Terms carefully and, if you agree to them, clicking on the boxes at the bottom of each of the two sets of provisions below to indicate that you understand, accept and will continually abide by these Terms. If in the future Databricks agrees to add any paid upgrade or other paid services to your Community Edition or Free Trial account, these Terms shall continue to apply in full unless you and Databricks mutually agree to be bound under a superseding set of governing terms and conditions.

YOUR DATA AND USE OF COMMUNITY EDITION – RESTRICTIONS APPLY

 **Databricks** billing@databricks.com [via](#) amazonses.com 2:55 PM (0 minutes ago) ☆  

to me ▾

 **databricks™**
making big data simple

Welcome to Databricks Community Edition!

Databricks Community Edition provides you with access to a free micro-cluster as well as a cluster manager and a notebook environment - ideal for developers, data scientists, data engineers and other IT professionals to get started with Spark.

We need you to verify your email address by clicking on [this link](#). You will then be redirected to Databricks Community Edition!

Get started by visiting: <https://accounts.cloud.databricks.com/signup/validate?emailToken=fa75685c08abf19c29c1fd1bf4f51f01&ce=true>

If you have any questions, please contact feedback@databricks.com.

- The Databricks Team



Sign In to Databricks



ENTER EMAIL ADDRESS

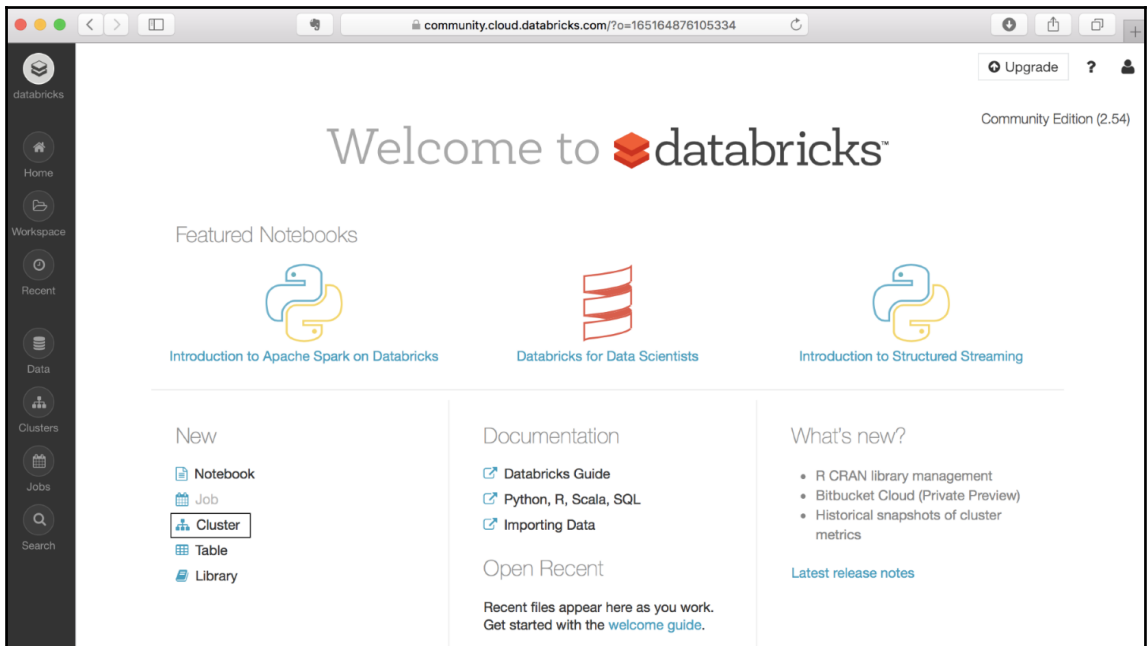


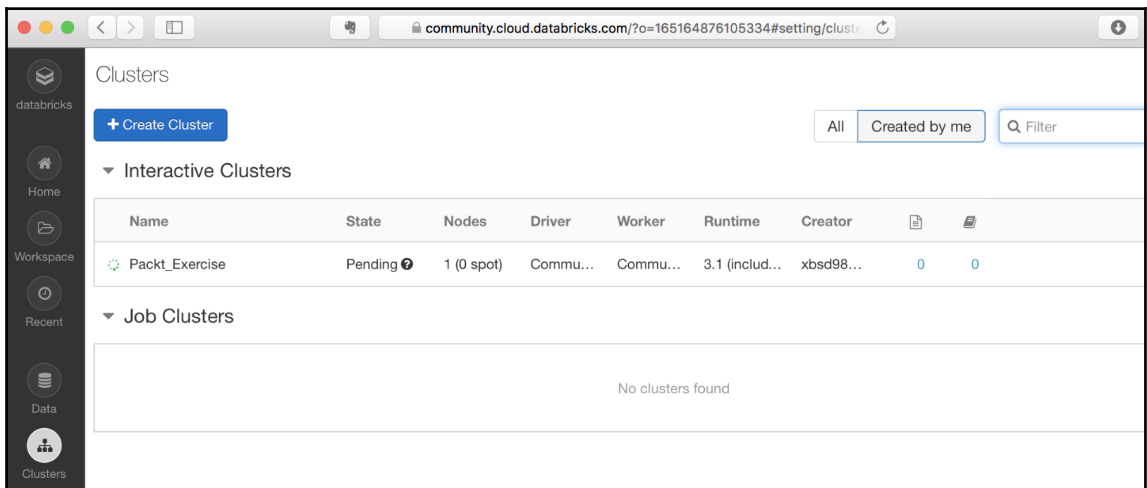
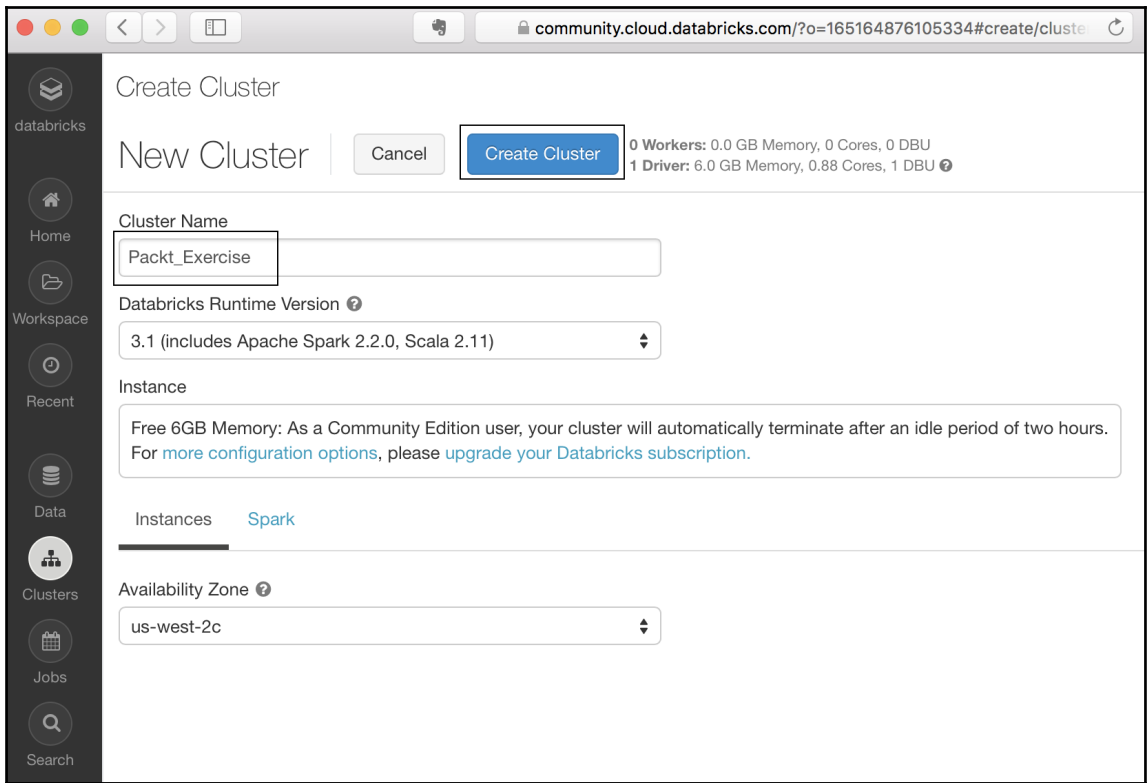
ENTER PASSWORD

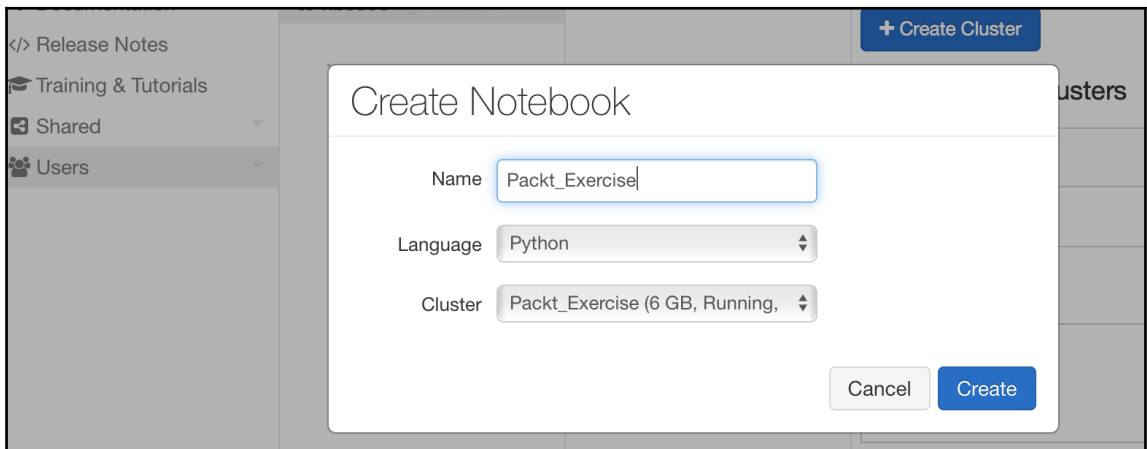
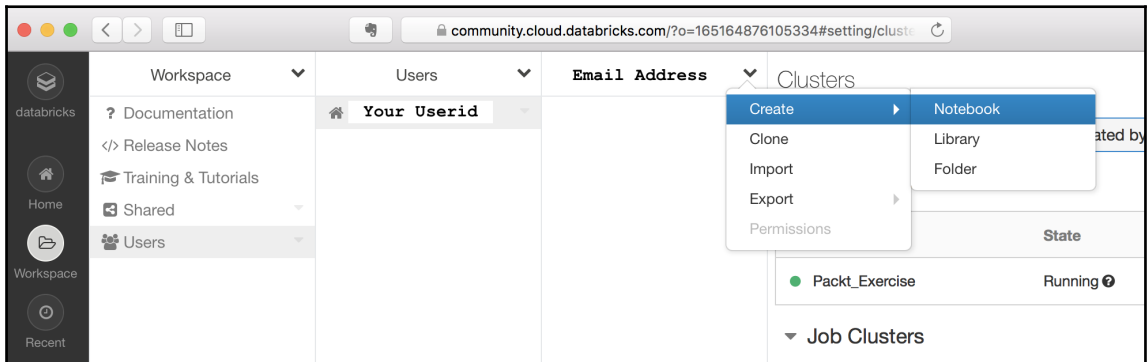
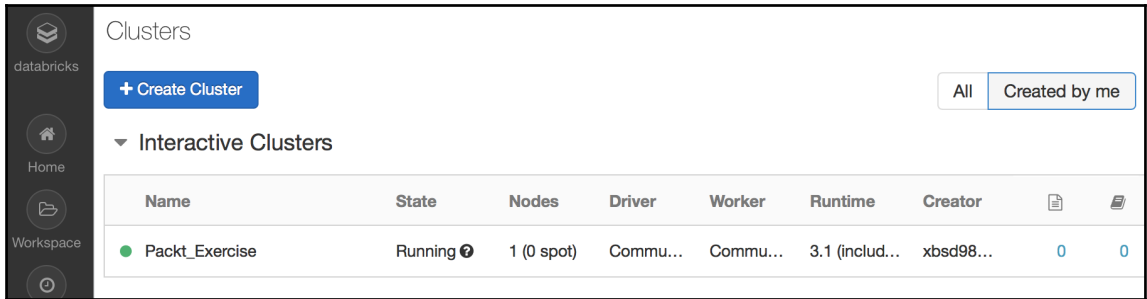
[Forgot Password?](#)

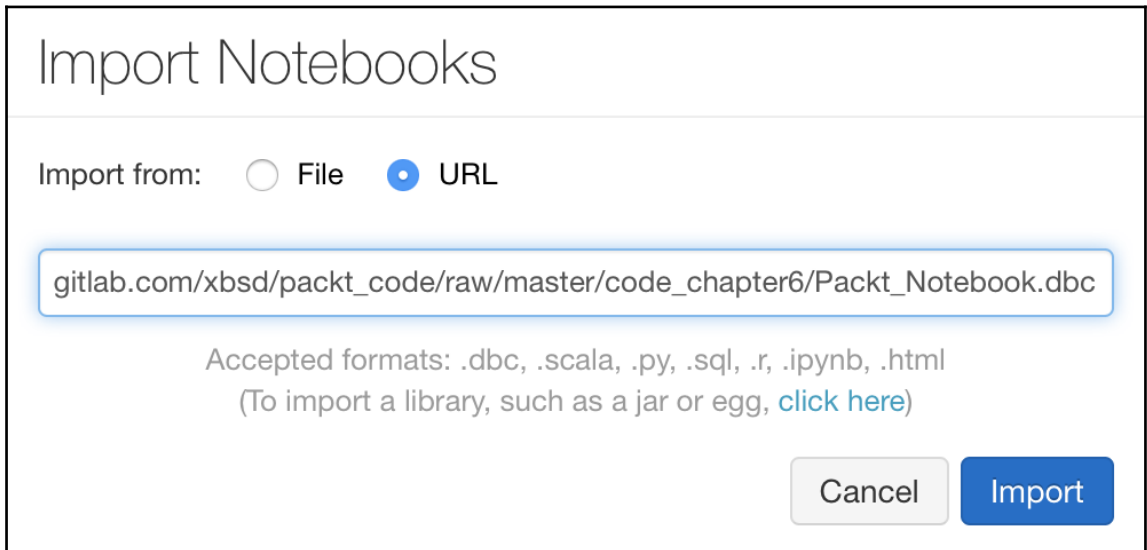
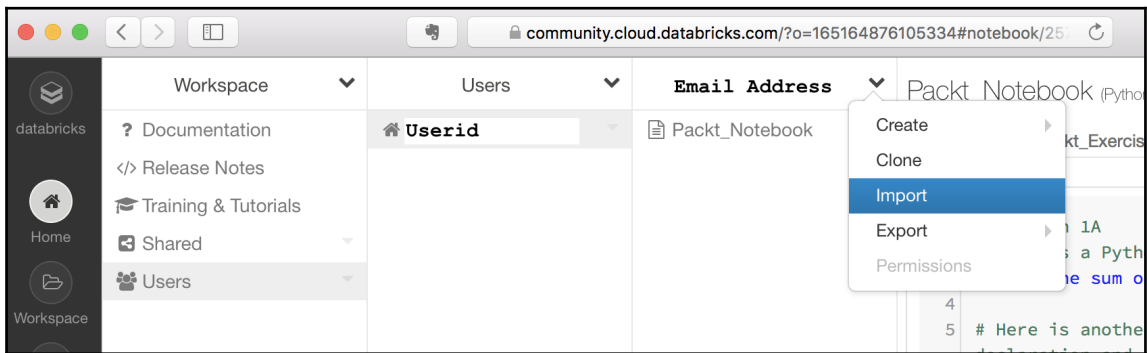
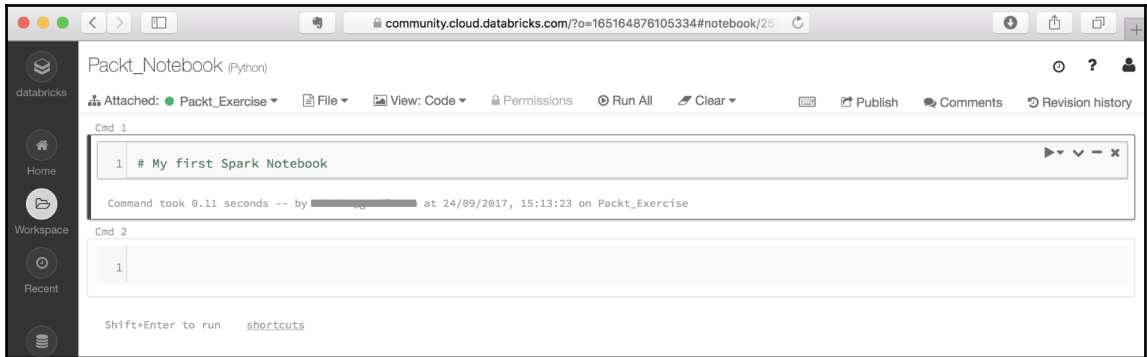
Sign In

New to Databricks? [Sign Up.](#)





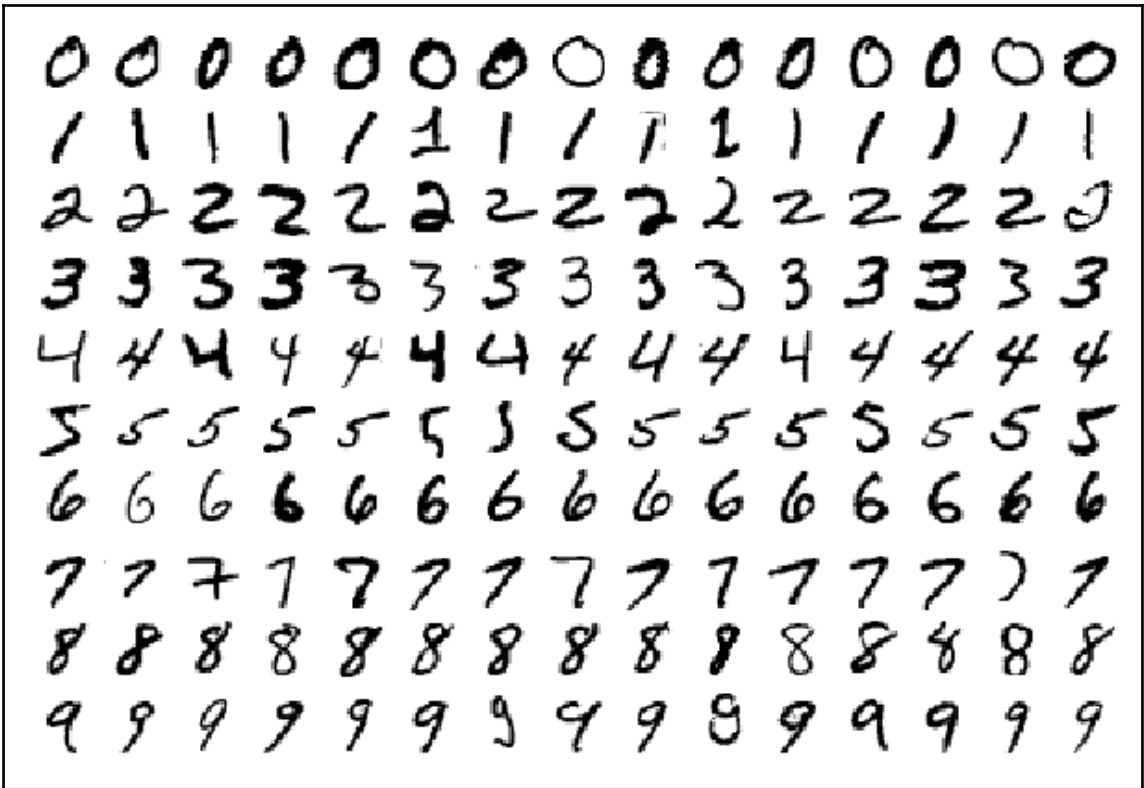




Workspace ▼	Users ▼	Email Address ▼
? Documentation	🏠 Userid ▼	📄 Packt_Notebook ▼

Chapter 7: An Introduction to Machine Learning Concepts

cyl	hp	wt	gear	mpg
6	110	2.620	4	21
6	110	2.875	4	21
4	93	2.320	4	22.8
6	110	3.215	3	21.4
8	175	3.440	3	18.7
6	105	3.460	3	MISSING
8	245	3.570	3	MISSING
4	62	3.190	4	MISSING
4	95	3.150	4	MISSING
6	123	3.440	4	MISSING





cyl	hp	wt	gear	mpg
6	110	2.620	4	MISSING
6	110	2.875	4	MISSING
4	93	2.320	4	MISSING
6	110	3.215	3	MISSING
8	175	3.440	3	MISSING
6	105	3.460	3	MISSING
8	245	3.570	3	MISSING
4	62	3.190	4	MISSING
4	95	3.150	4	MISSING
6	123	3.440	4	MISSING

The image shows a Google search interface. At the top left is the Google logo. The search bar contains the text "images of Saturn rings from Cassini" and a magnifying glass icon. Below the search bar are navigation tabs: "All", "Images", "News" (which is selected and underlined), "Videos", "Shopping", "More", "Settings", and "Tools".

Below the tabs, it says "About 31,800 results (0.84 seconds)".

The main content area features a large image of Saturn's rings on the left. To its right is a news snippet with the following text:

- Surprising Science from Cassini's Grand Finale**
- Sky & Telescope - Oct 17, 2017
- You'd think scientists would have Saturn all figured out after watching it up close for 13 ... Cassini took this **image of Saturn's rings** on Sept.
- Get lost in NASA's stunning Saturn rings panorama**
- CNET - Oct 17, 2017
- Fresh findings from Cassini**
- Phys.Org - Oct 17, 2017

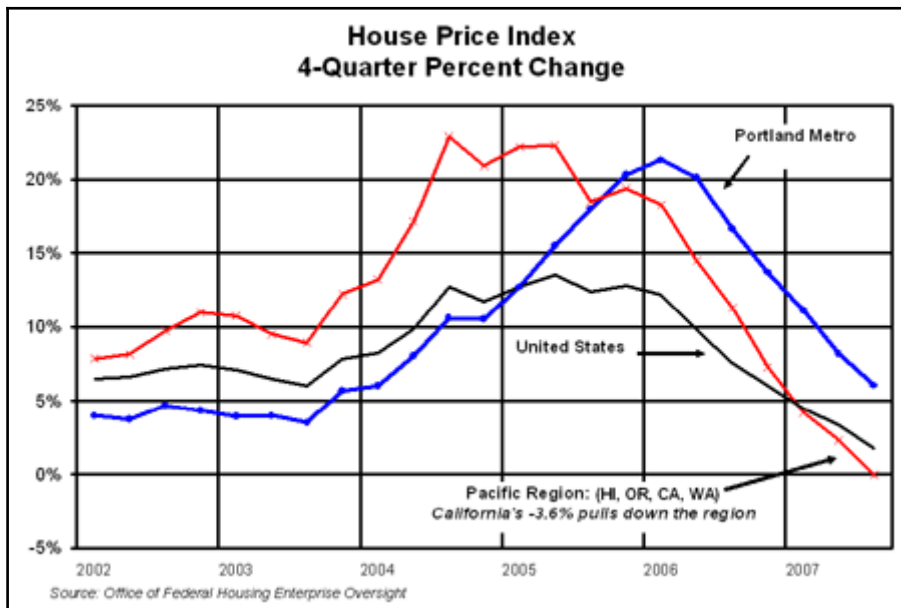
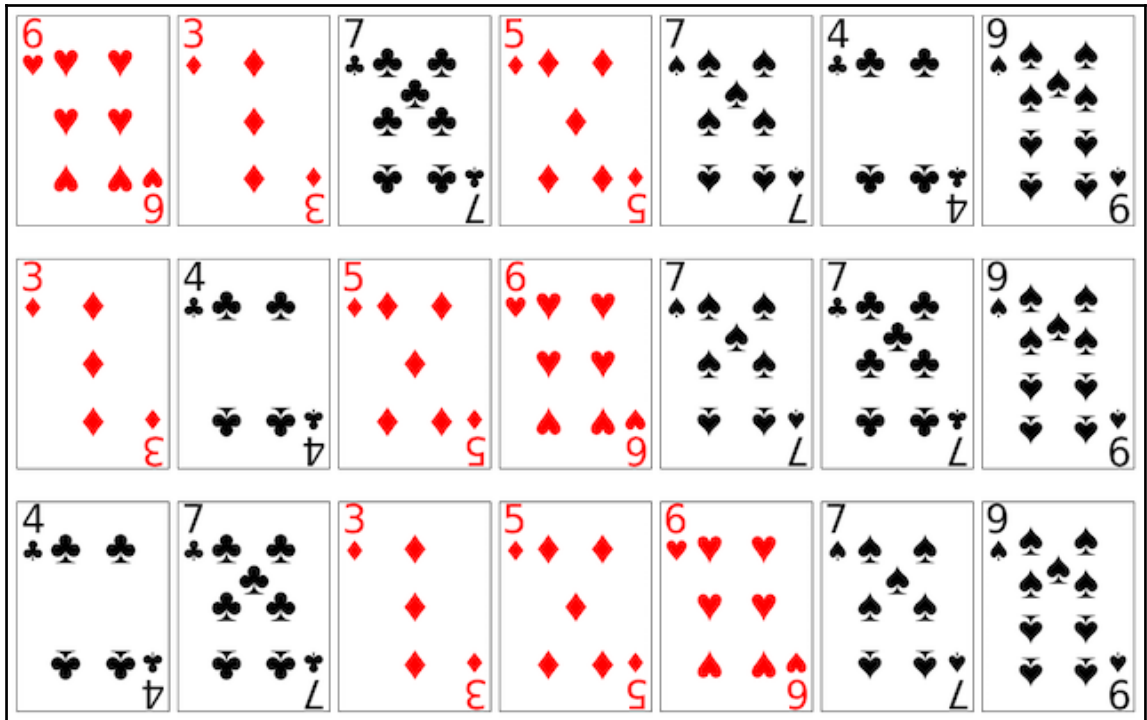
Below this snippet are several other news items:

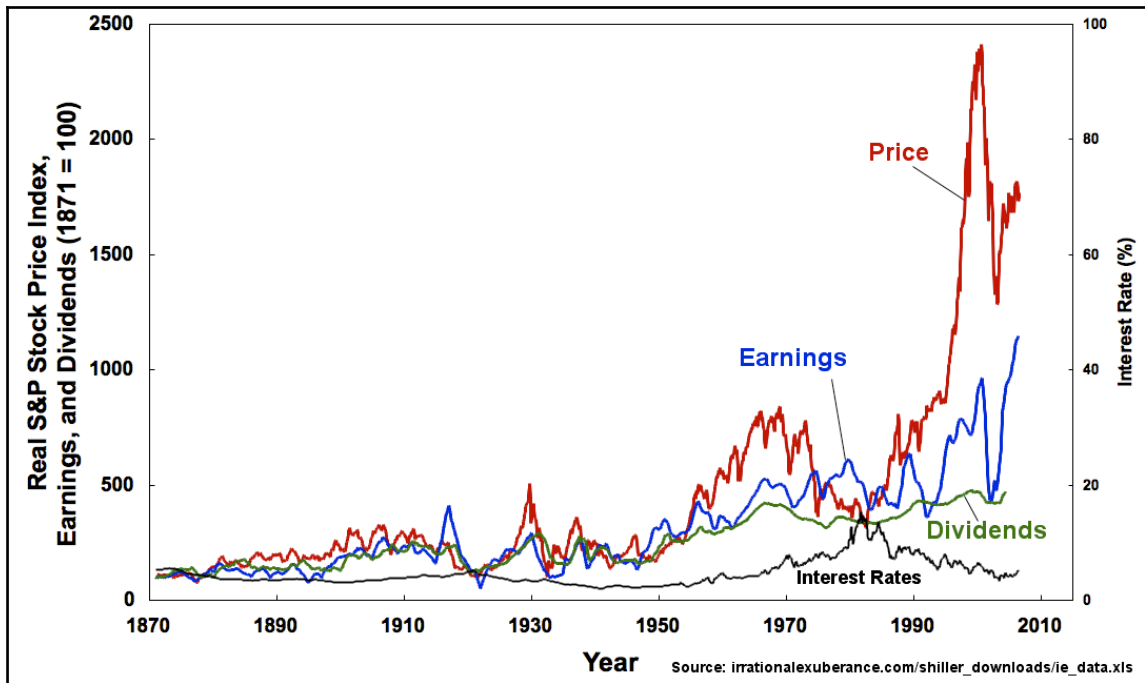
- Moons work together to hold Saturn's largest ring in place**
- Futurity: Research News - Oct 17, 2017
- Cassini data finds that Saturn's largest ring is kept in place by 7 of its ...**
- The Indian Express - 10 hours ago
- Cassini's death dive discovers a 'confederation of moons' hold ...**
- In-Depth - Daily Mail - Oct 17, 2017

At the bottom of the news section, there is a row of six small thumbnail images of Saturn's rings. Below each thumbnail is a label: "CNET", "Daily Mail", "Phys.Org", "Futurity: Res...", "The Indian Ex...", and "Science Daily".

At the bottom left of the news section, there is a red-bordered button with the text "View all".



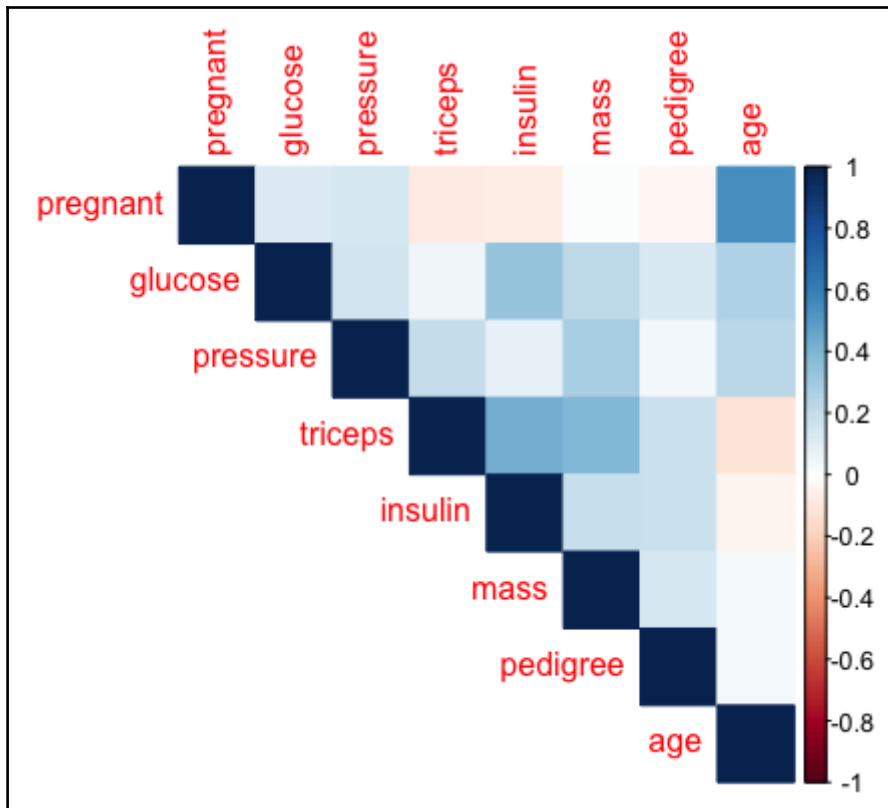


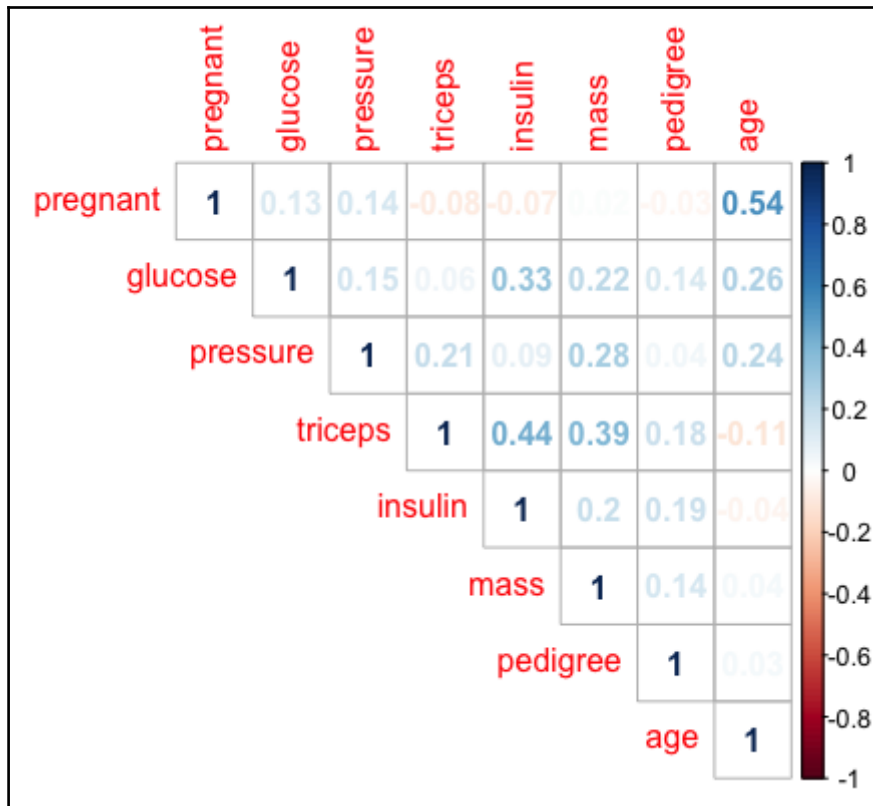


date	time	agency	type	prepaid	rate	amount
8/5/17	1:05 PM	ABTA	TOLL	Y	STANDARD	\$0.78
8/12/17	5:00 PM	NCTA	TOLL	Y	STANDARD	\$0.43
8/12/17	4:39 PM	NCTA	TOLL	Y	STANDARD	\$0.43
8/13/17	4:00 PM	ABTA	PASS	Y	STANDARD	\$0.43
8/14/17	1:52 AM	DSTC	TOLL	Y	STANDARD	\$0.43
8/14/17	3:45 PM	NCTA	PASS	Y	STANDARD	\$0.78
8/15/17	7:37 AM	DSTC	TOLL	Y	STANDARD	\$0.78
8/17/17	8:42 PM	NCTA	TOLL	Y	STANDARD	\$0.98

Concept	Terminology	Meaning
Type of Variable	Continuous/Quantitative	Numbers on which you can perform arithmetic
	Discrete/Categorical	Qualitative – Alphabetical, or Numeric that simply denotes a 'category'
Variable in an equation	Dependent	The Left-Hand-Side, the y in $y = x + 1$
	Independent	The Right-Hand-Side, every variable other than the y in $y = x + 1$ (i.e., x in this case)
Type of ML	Regression	You want to predict a numeric outcome
	Classification	You want to predict a categorical outcome

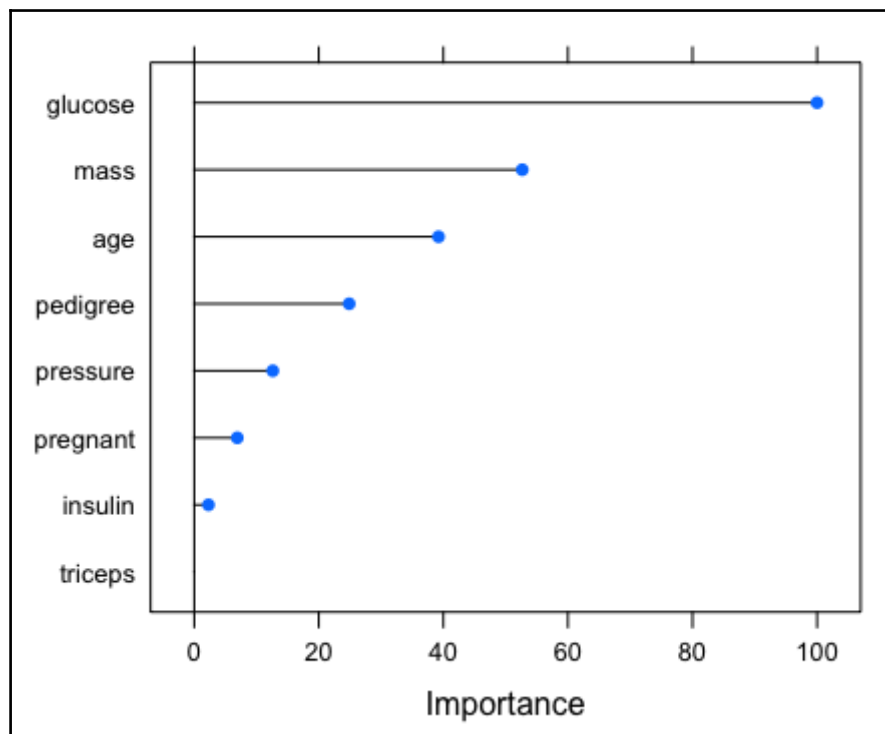
Scores	45	66	66	55	55	52	61	64	65	49
Scores - Mean	-12.8	8.2	8.2	-2.8	-2.8	-5.8	3.2	6.2	7.2	-8.8
Divide by SD	-1.67	1.07	1.07	-0.4	-0.4	-0.8	0.42	0.81	0.94	-1.2



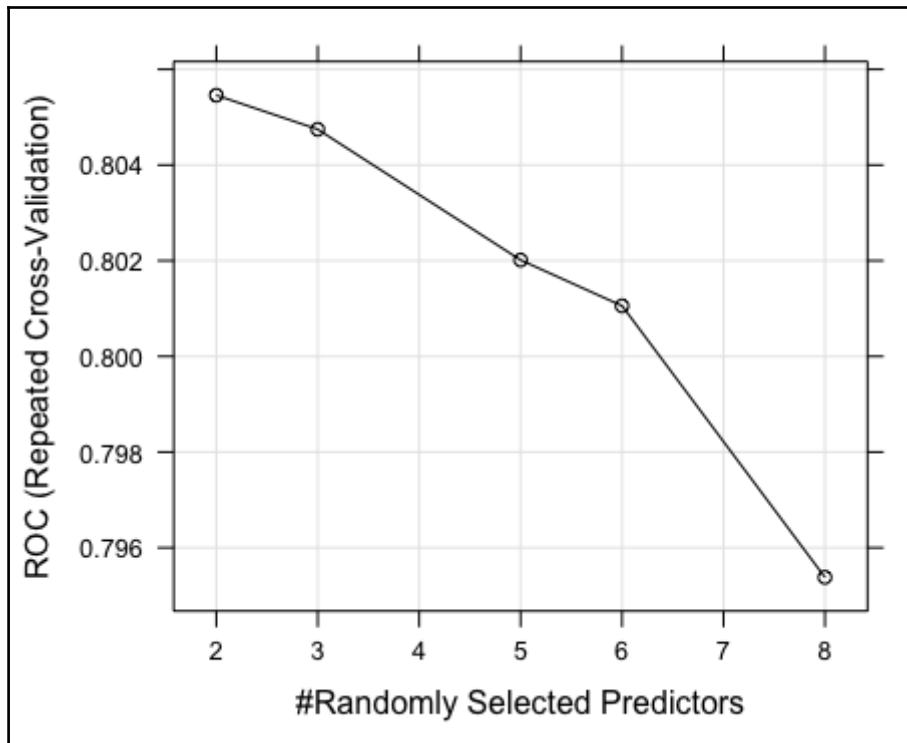


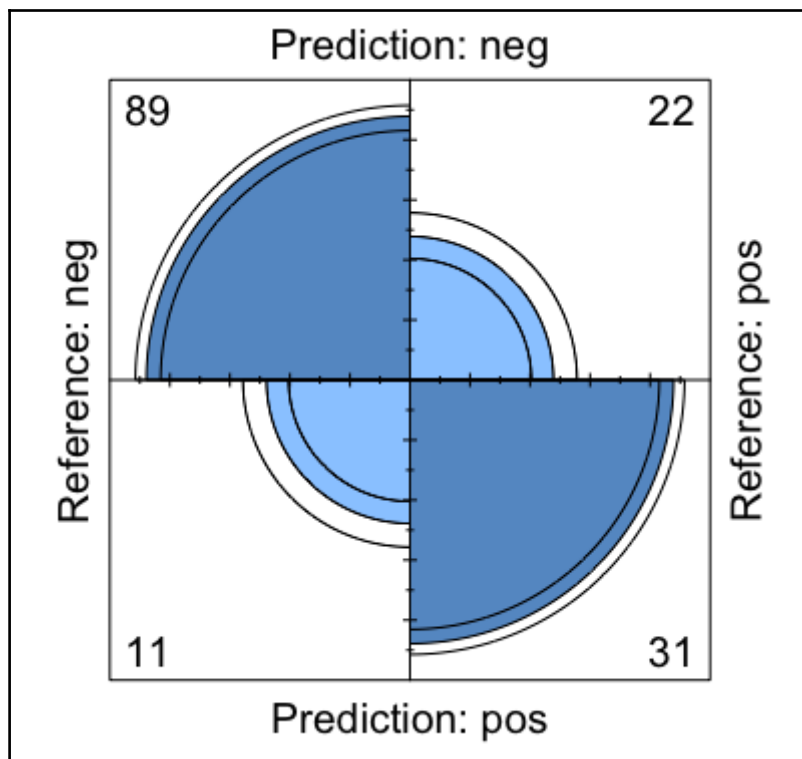
	pregnant	glucose	pressure	triceps	insulin	mass	pedigree	age	diabetes
5	0	137	40	35	168	43.1	2.288	33	pos
6	5	116	74	0	0	25.6	0.201	30	neg
7	3	78	50	32	88	31.0	0.248	26	pos
8	10	115	0	0	0	35.3	0.134	29	neg
9	2	197	70	45	543	30.5	0.158	53	pos
10	8	125	96	0	0	NA	0.232	54	pos
11	4	110	92	0	0	NA	0.191	30	neg
12	10	168	74	0	0	NA	0.537	34	pos
13	10	139	80	0	0	NA	1.441	57	neg
14	1	189	60	23	846	NA	0.398	59	pos
15	5	166	72	19	175	NA	0.587	51	pos

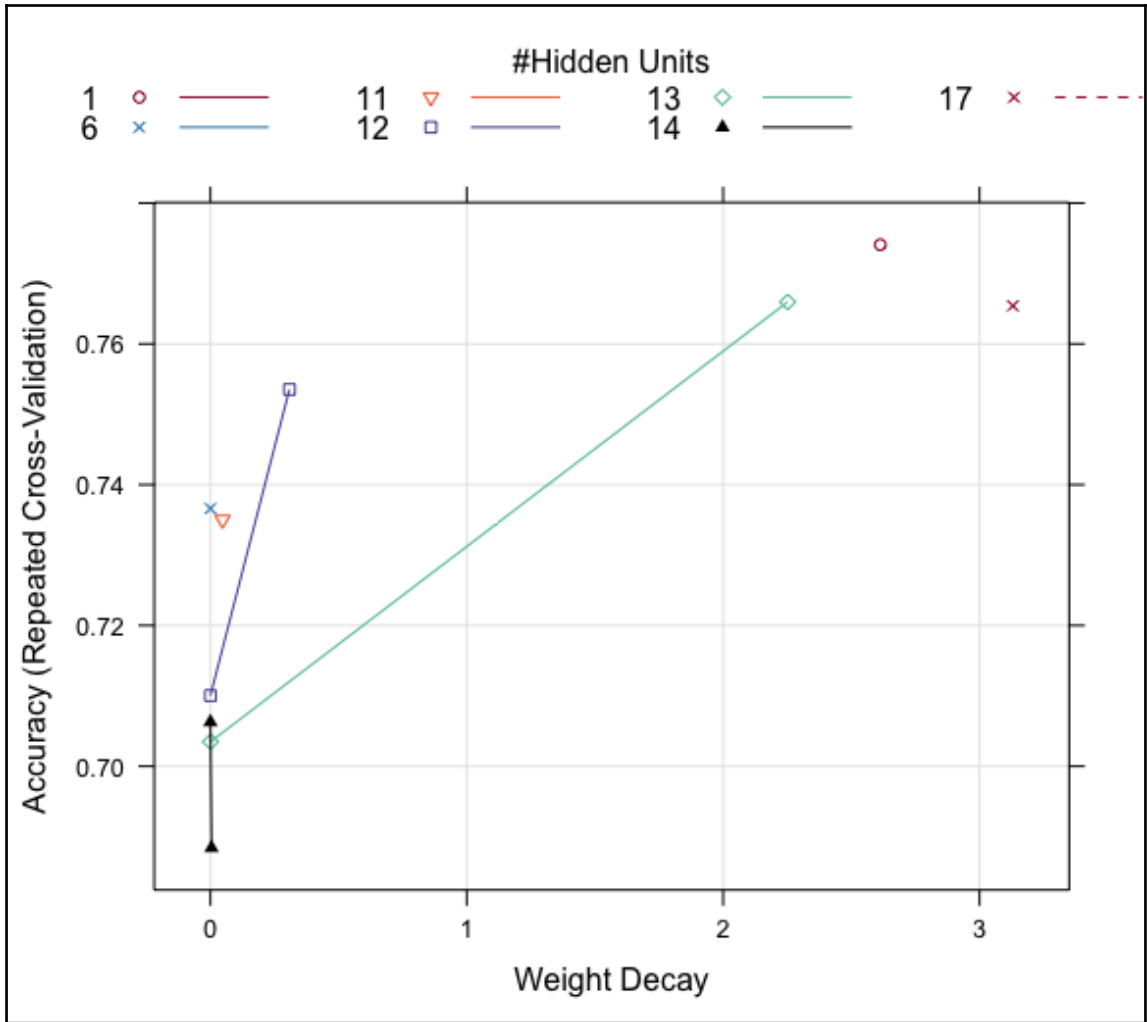
```
> data.frame(actual=actual[10:100], impute_with_mean=diabmiss$mass[10:100])
  actual impute_with_mean
1    0.0          32.12378
2   37.6          32.12378
3   38.0          32.12378
4   27.1          32.12378
5   30.1          32.12378
6   25.8          32.12378
7   30.0          32.12378
8   45.8          32.12378
```

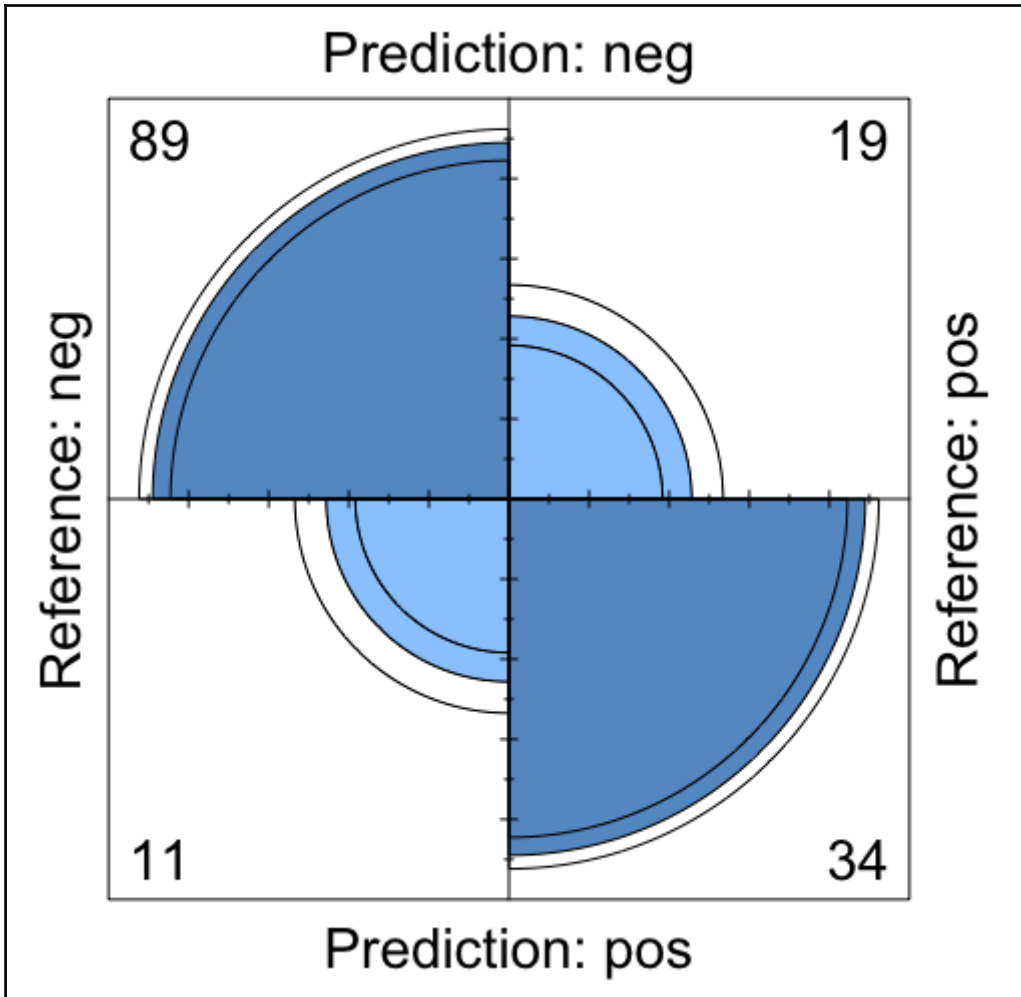


METHOD 1: <u>WITHOUT</u> CROSS-VALIDATION	METHOD 2: <u>WITH</u> CROSS-VALIDATION
<div style="border: 1px solid black; margin-bottom: 10px; padding: 5px; display: flex; justify-content: space-around; background-color: #e6f2ff;"> TRAINING SET TEST SET </div> <ol style="list-style-type: none"> 1. SPLIT DATA INTO 70-30 OR 80-20 2. SET ASIDE THE 20% PART FOR TESTING 3. CREATE MODEL WITH DIFFERENT MODEL PARAMETERS USING THE TRAINING SET 4. APPLY THE FINAL MODEL USING THE SELECTED PARAMETERS ON THE TEST SET 5. GET ACCURACY AND PERFORMANCE ESTIMATES 	<p style="text-align: center;">A 5-FOLD CROSS-VALIDATION</p> <div style="border: 1px solid black; margin-bottom: 10px; padding: 5px; display: flex; justify-content: space-around; background-color: #e6f2ff;"> FOLD 1 FOLD 2 FOLD 3 FOLD 4 FOLD 5 TEST SET </div> <ol style="list-style-type: none"> 1. SPLIT DATA INTO 70-30 OR 80-20 2. SET ASIDE THE 20% PART FOR TESTING 3. SPLIT THE TRAINING DATASET INTO k-FOLDS 3. CREATE MODEL WITH DIFFERENT MODEL PARAMETERS ON k - 1 FOLDS AND TEST THE PERFORMANCE ON THE FOLD THAT WASN'T SELECTED 4. REPEAT STEP 3 FOR EACH OF THE k FOLDS 5. IDENTIFY THE PARAMETERS THAT GIVE THE BEST OVERALL PERFORMANCE 4. APPLY THE FINAL MODEL USING THE SELECTED PARAMETERS ON THE TEST SET 5. GET ACCURACY AND PERFORMANCE ESTIMATES

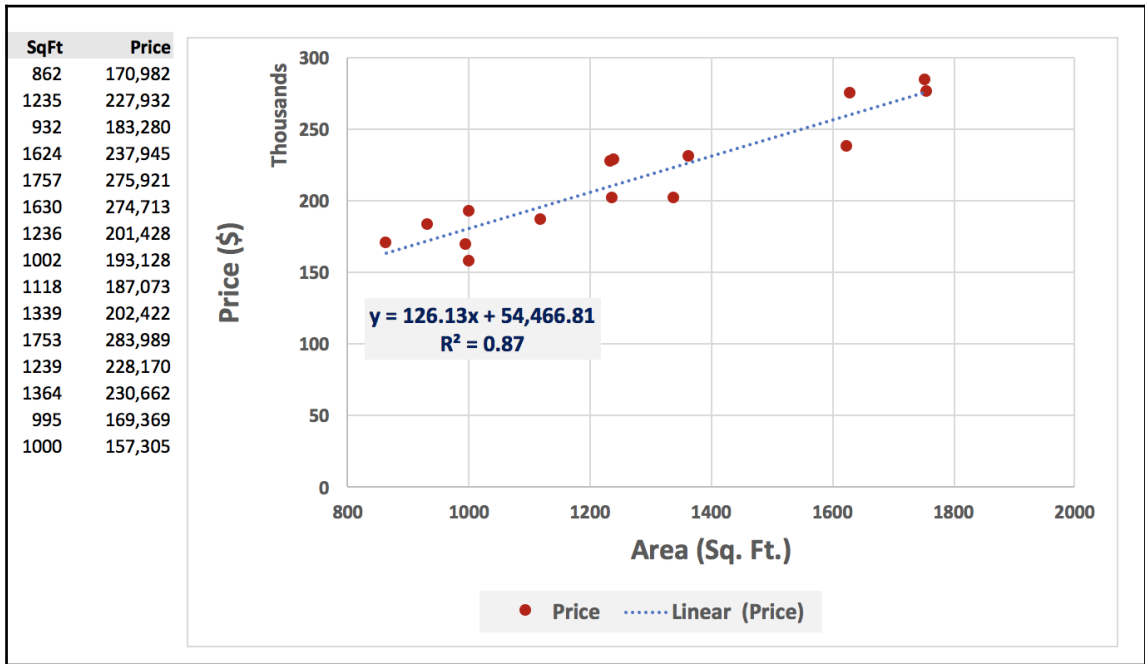


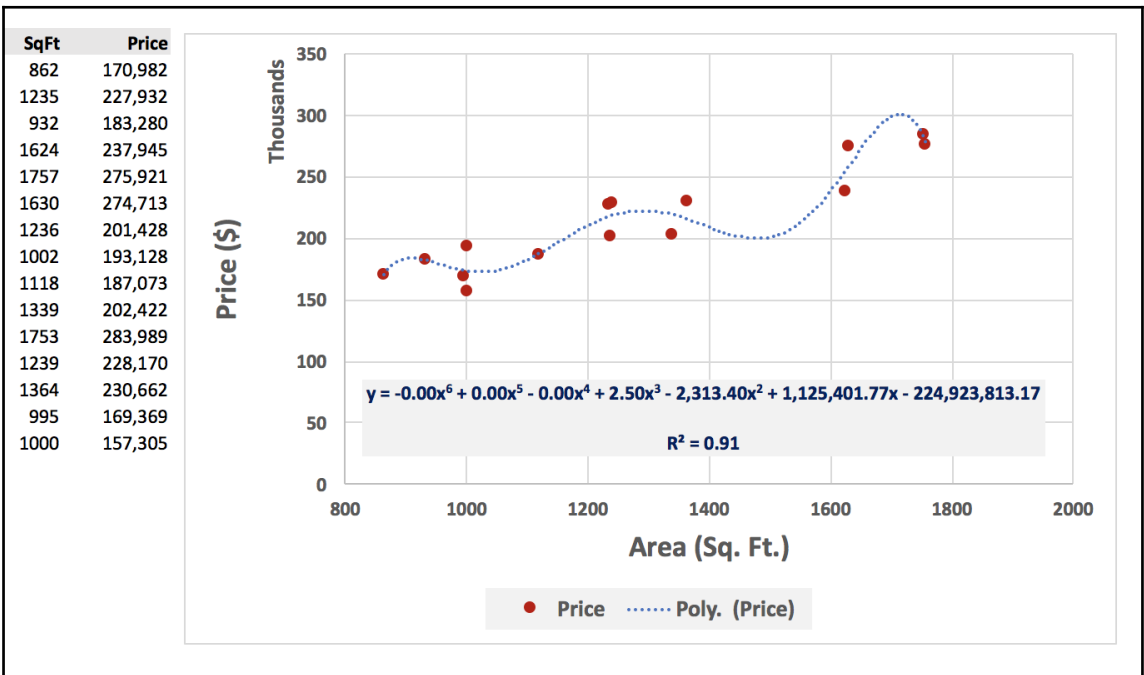
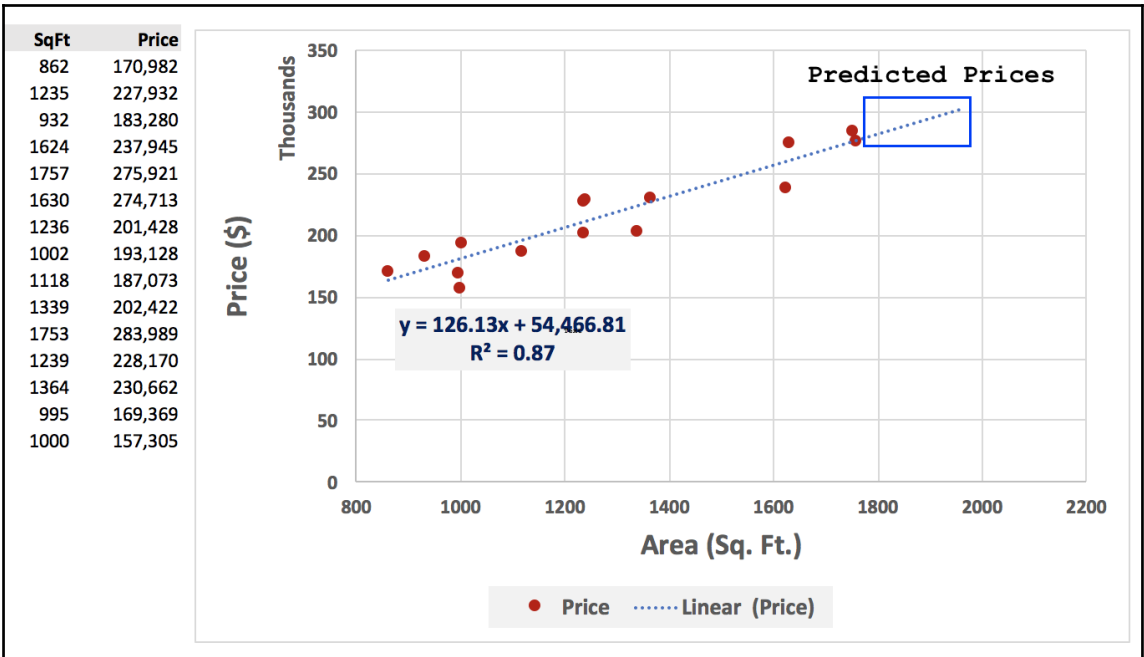


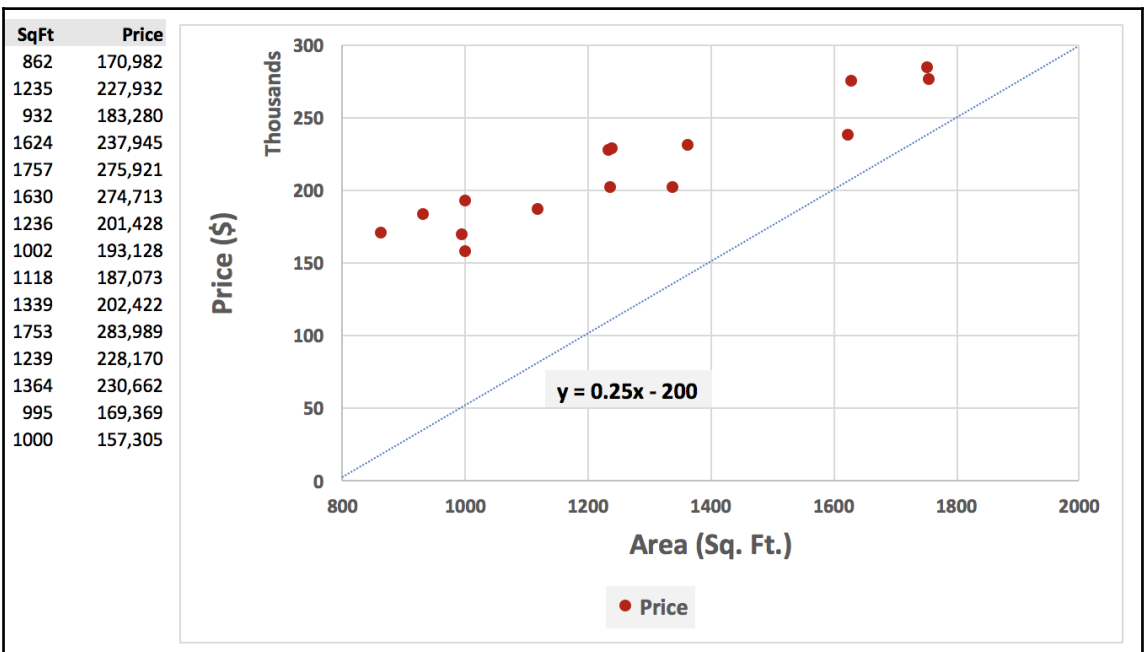
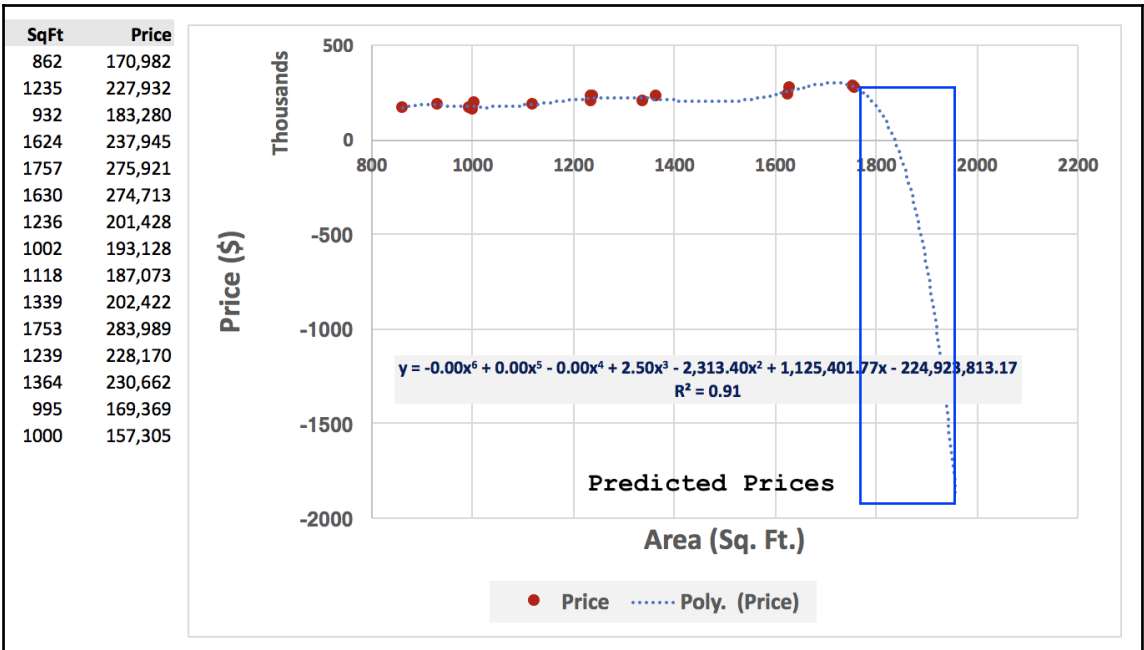




Chapter 8: Machine Learning Deep Dive



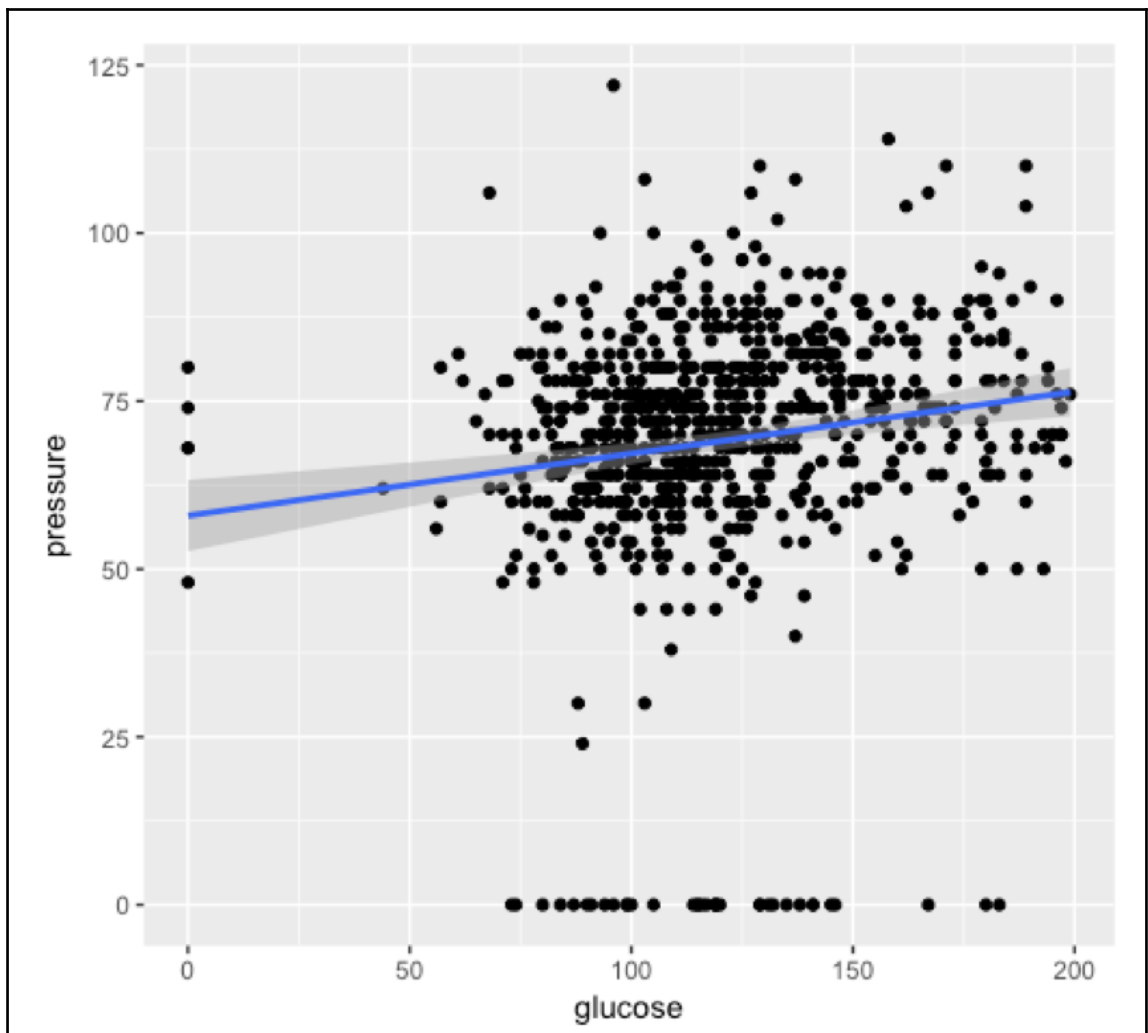


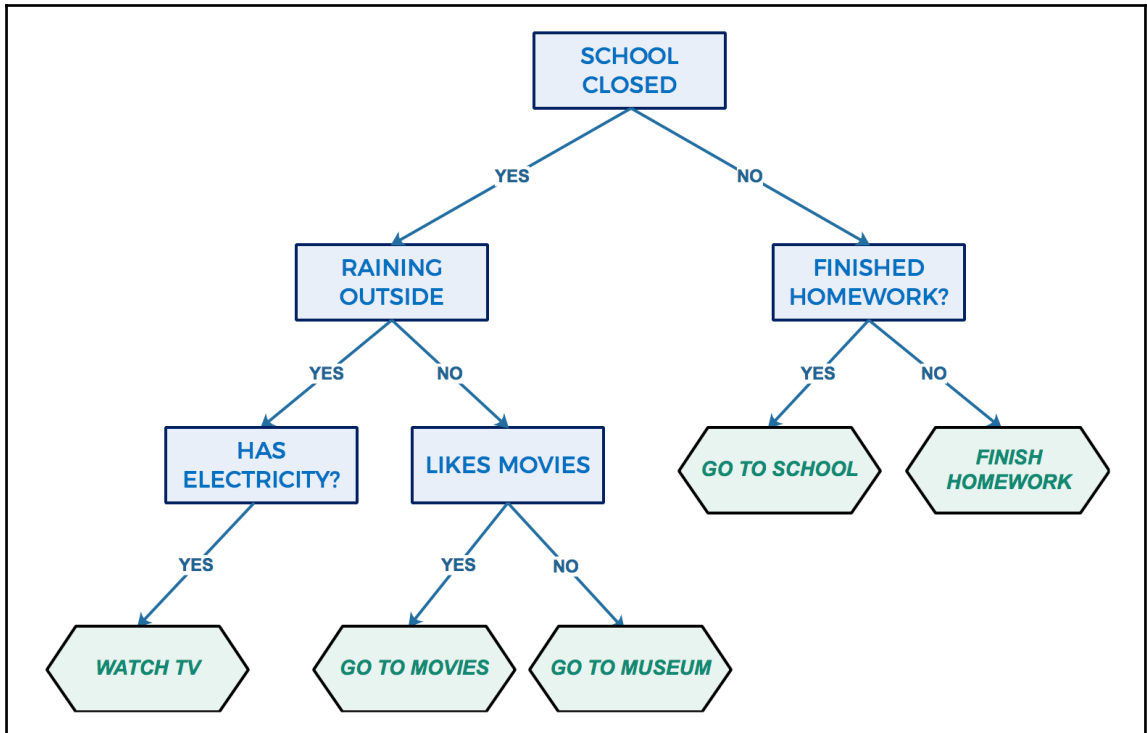


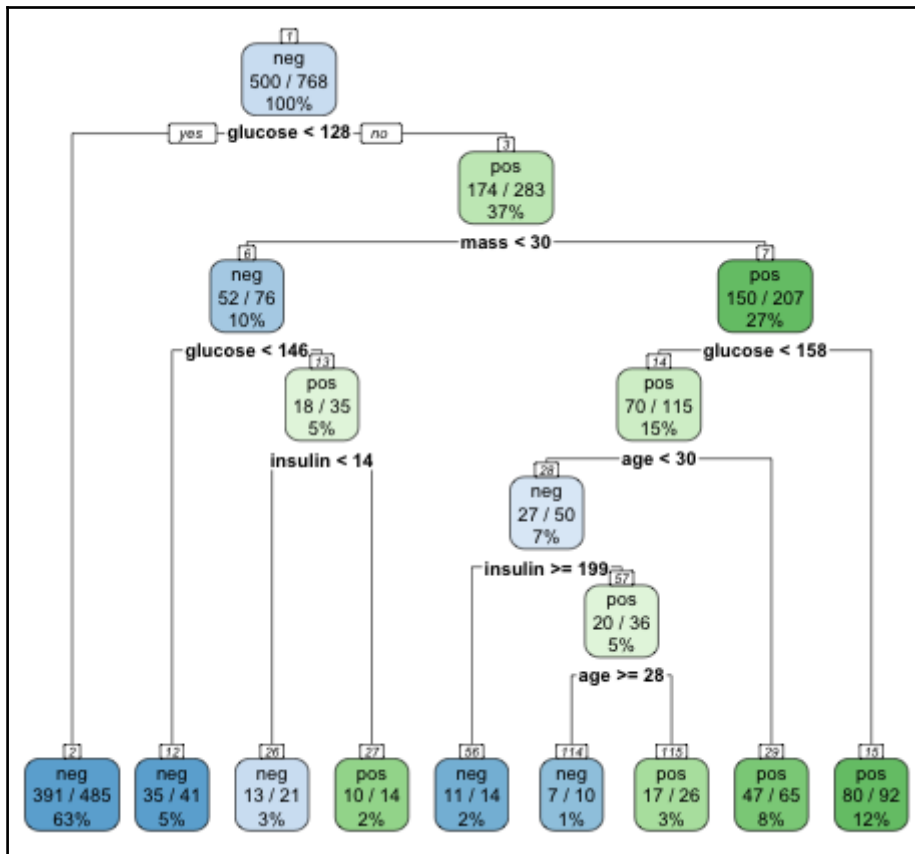


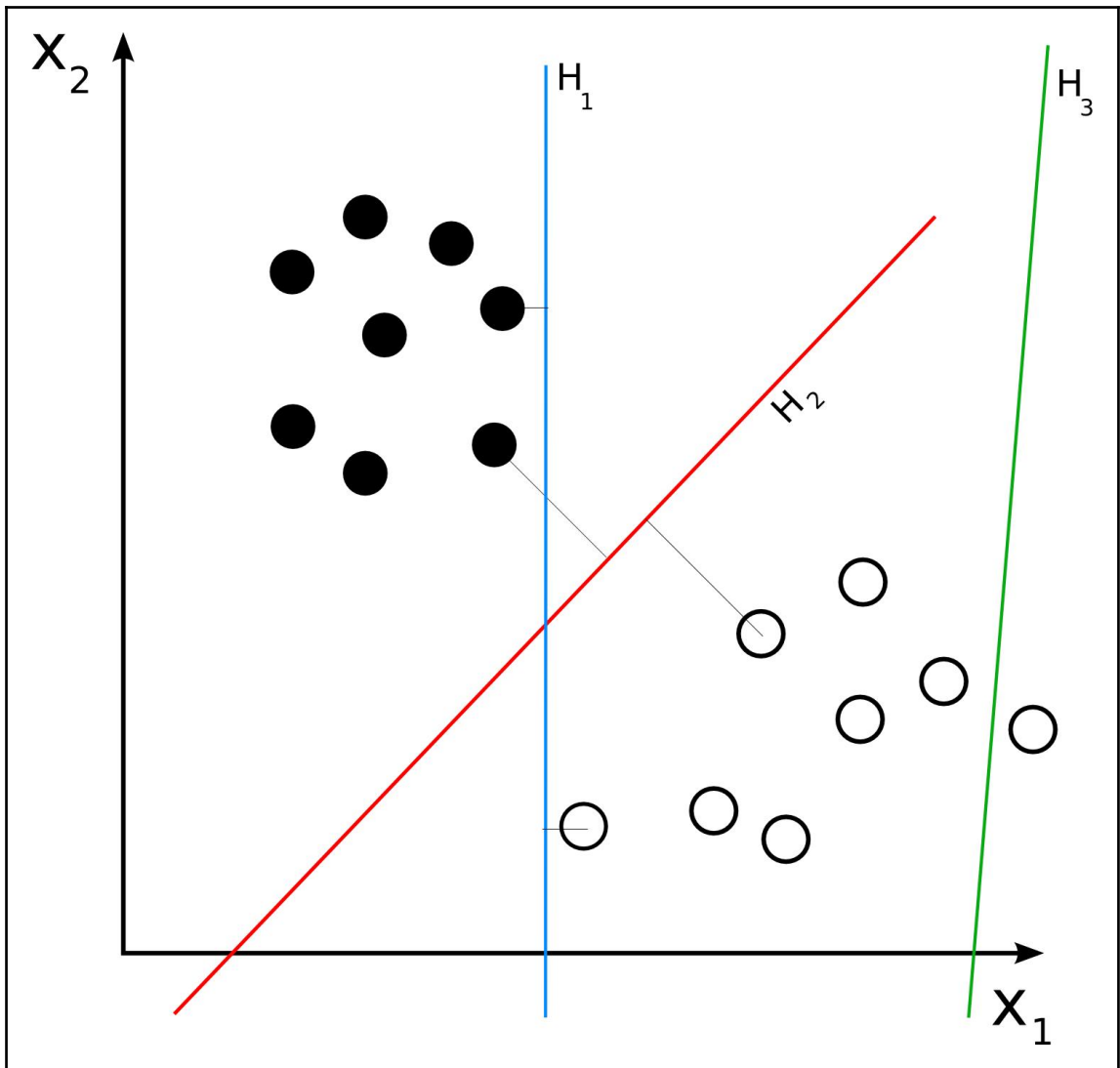


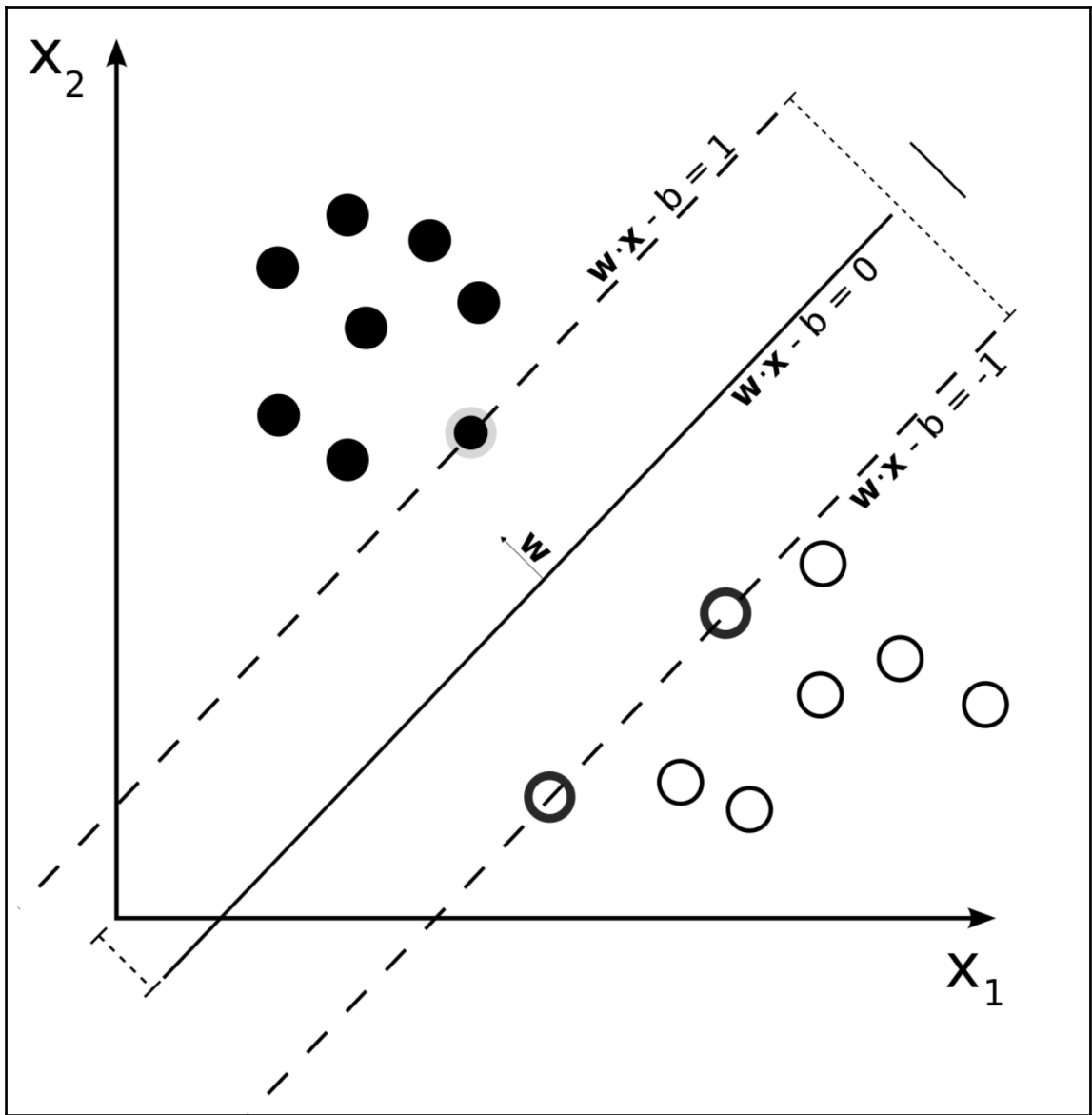
To Fix	The Issue	Remedies
High Variance <i>Unable to 'Generalise'</i>	Not enough information	Get more training examples
	Too many features (variables) are being considered	Reduce number of features
	Too less penalisation	Increase Regularisation parameter (lambda)
High Bias <i>We have generalised way too much</i>	Not enough features to build a more thorough model	Add more number of features
	Model is too simple	Add more complex features (eg., polynomial features)
	Too much penalisation	Decrease Regularisation parameter (lambda)

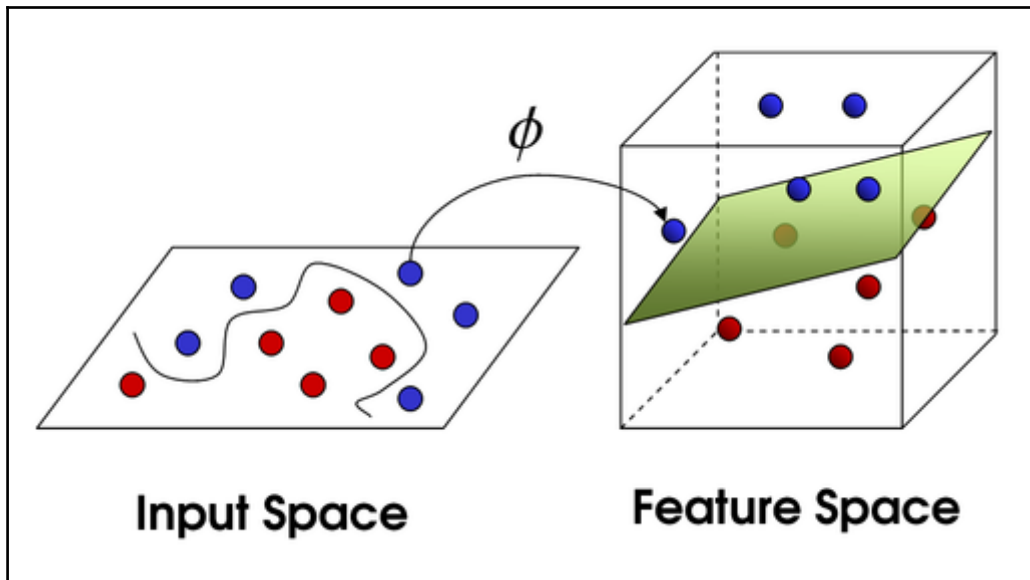


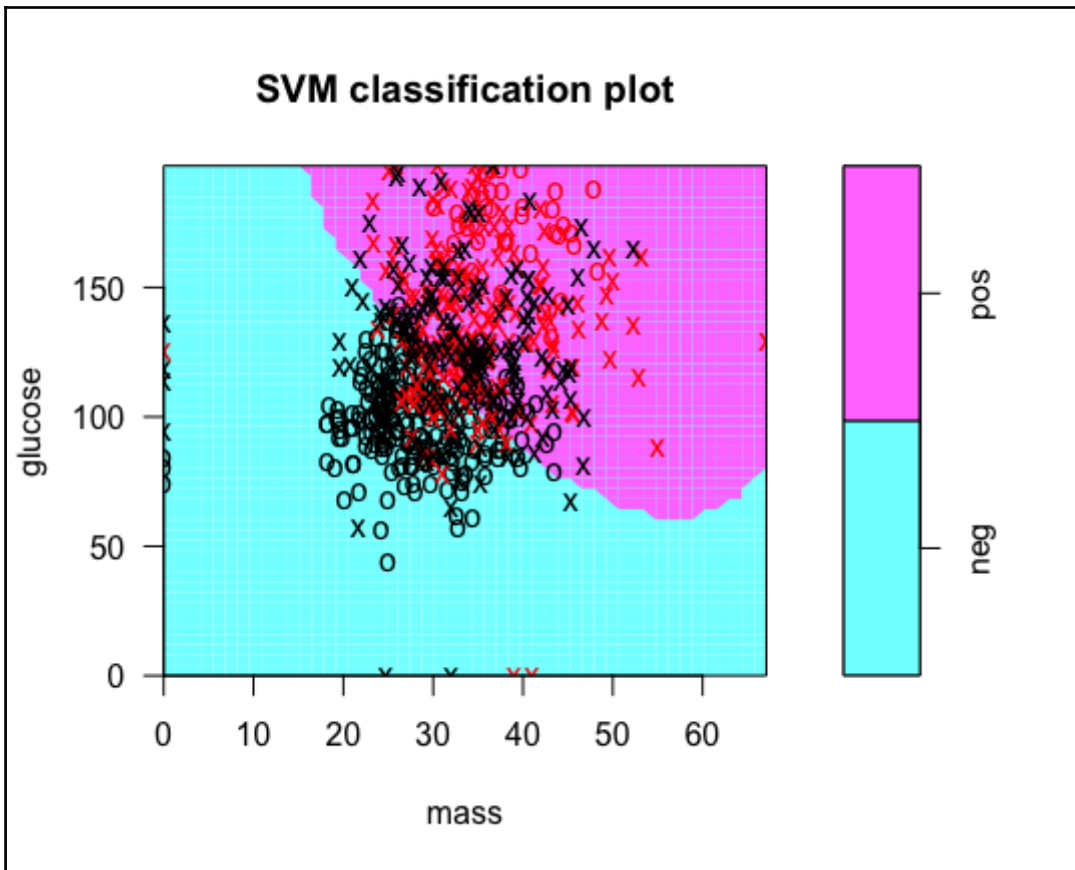


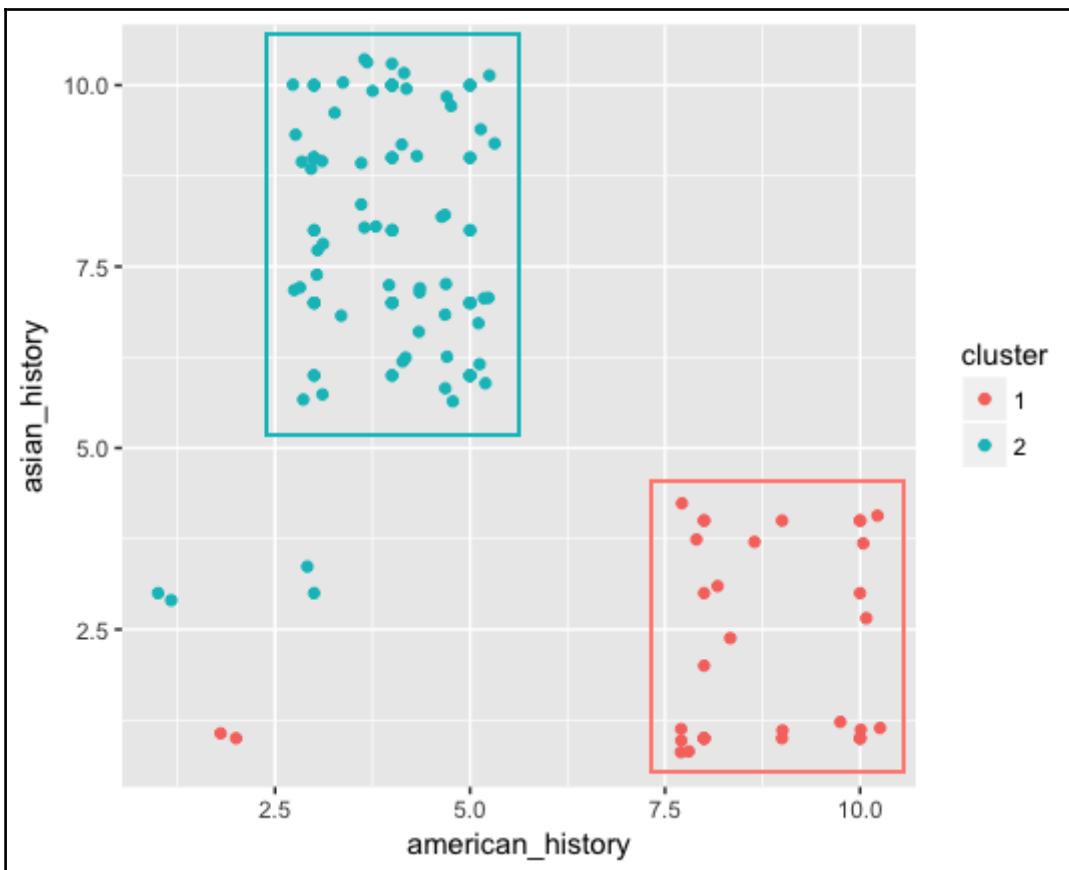


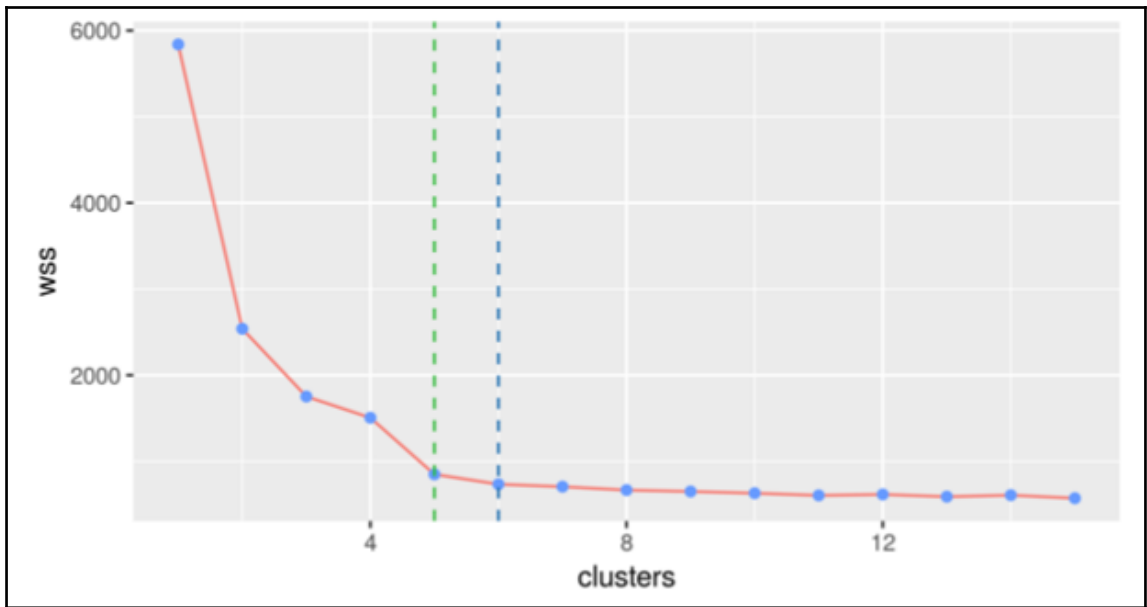


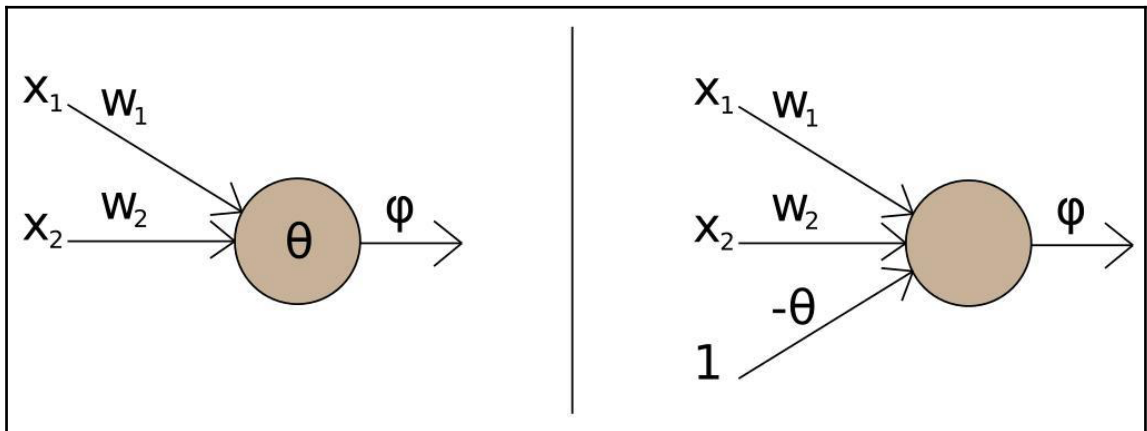
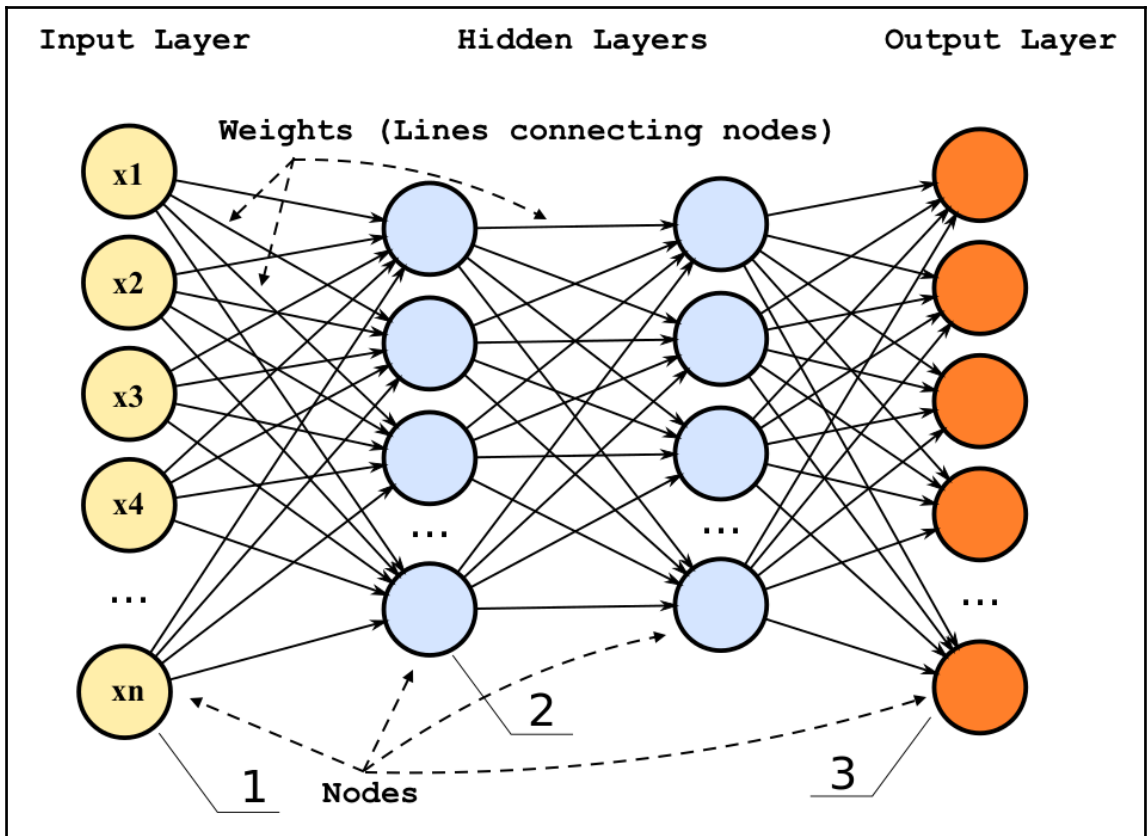




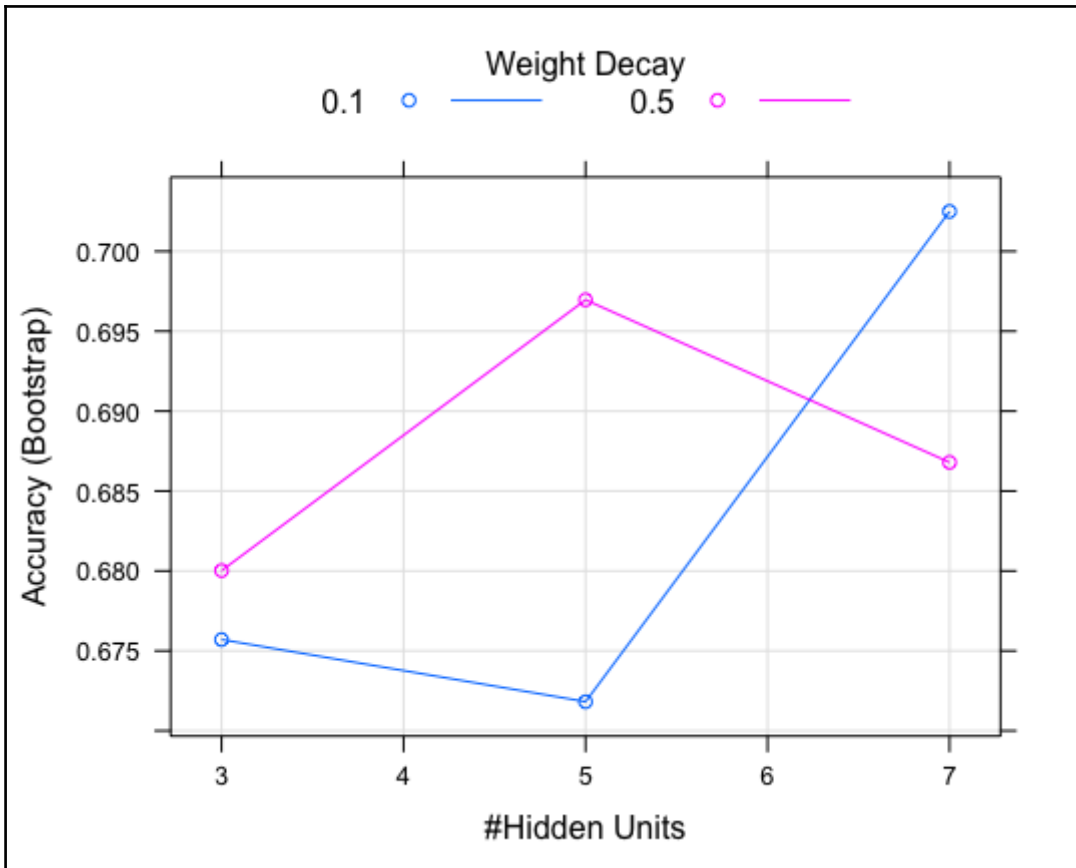








$$f\left(\sum_i (w_i * x_i) + b\right)$$



OpenPaymentsData.CMS.gov

Find Your Doctor's Payments

First Name Last Name Search

Explore Open Payments Data

Looking for a teaching hospital or company making payments?
For a more refined search, use the Search Tool.

- How to use the Search Tool
- Check out the new Payments by State page

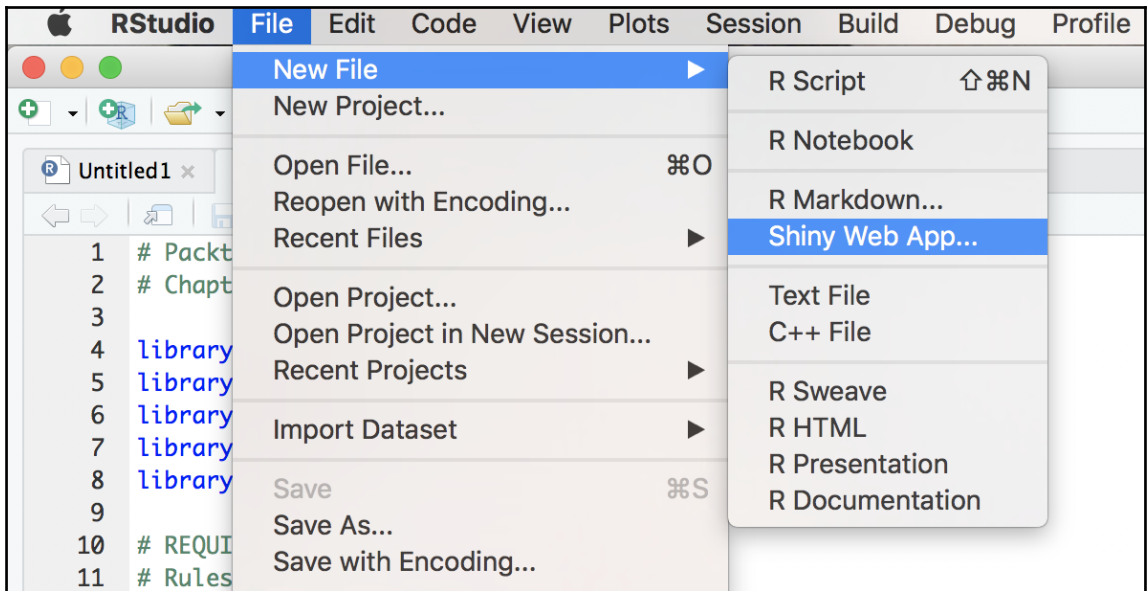
The Facts on Open Payments

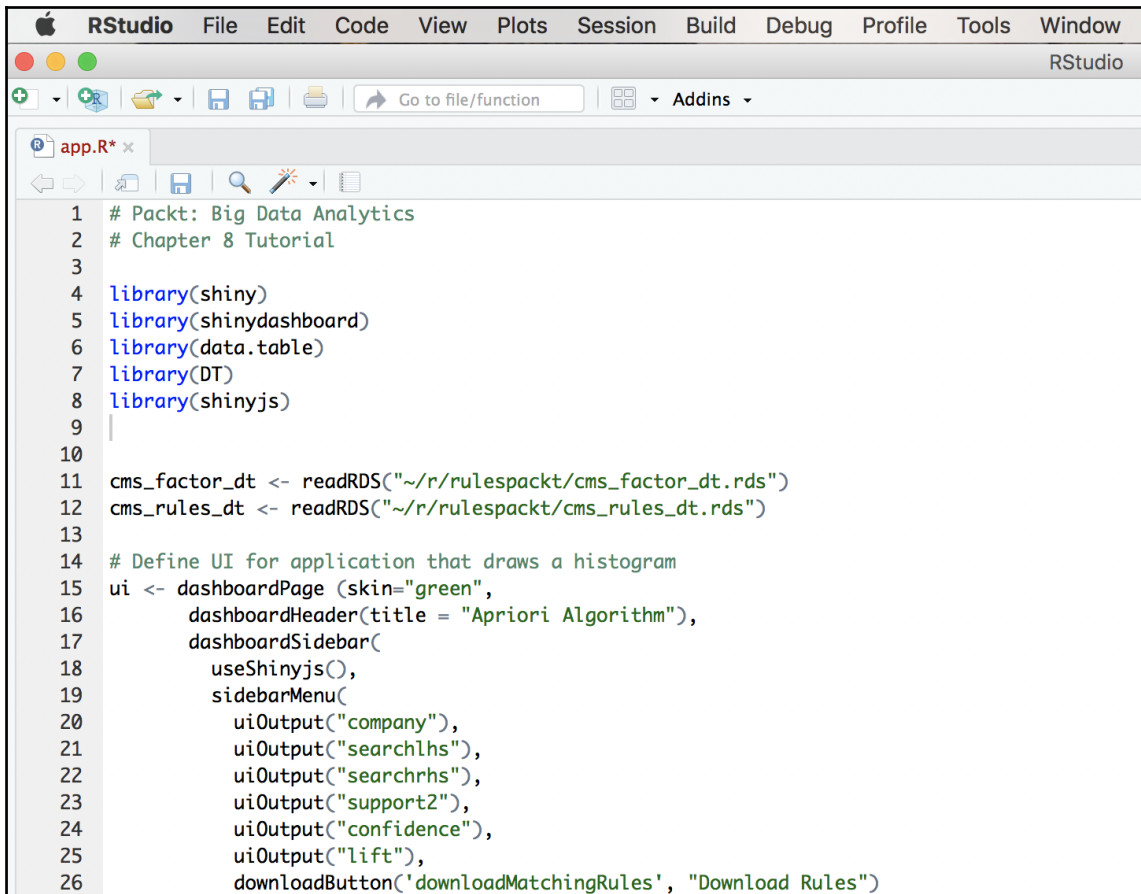
2013 Totals 2014 Totals 2015 Totals 2016 Totals Totals All Years

```

"category","city","company","firstName","lastName","payment","paymentNature","product","state"
"CARDIOVASCULAR AND METABOLISM","White Plains","AstraZeneca Pharmaceuticals LP","Ronald","Weissman","14.42","Food and Beverage","FARKIGA","NY"
"CARDIOVASCULAR","MIDDLETOWN","Novartis Pharmaceuticals Corporation","BONNIE","SEECHARRAN","4.58","Food and Beverage","ENTRESTO","NY"
"Immunology","ALBANY","Janssen Pharmaceuticals, Inc","DOROTA","HAUSNER-SYPEK","23.18","Food and Beverage","SIMPONI ARIA","NY"
"RAYOS","BROOKLYN","Horizon Pharma plc","PLACIDO","MORANO","5.4","Food and Beverage","RAYOS","NY"
"SCS IPGs","NEW YORK","St. Jude Medical, Inc","Alon","Mogilner","49.05","Food and Beverage","Eon and EonC and Eon Mini","NY"
"PSYCHIATRY","STATEN ISLAND","Otsuka America Pharmaceutical, Inc.,"SHAKEEB","HUSSAIN","14.07","Food and Beverage","ABILIFY MAINTENA","NY"
"Optics","MIDDLE VILLAGE","Abbott Laboratories","NERRY","NAYZOV","96.38","Food and Beverage","Tecnis 1-piece IOL","NY"
"RESPIRATORY","Rochester","AstraZeneca Pharmaceuticals LP","Gary","Glazer","14.96","Food and Beverage","SYMBICORT","NY"
"CARDIOVASCULAR","BRONX","Novartis Pharmaceuticals Corporation","FERNANDO","MORA-MCLAUGHLIN","20.04","Food and Beverage","ENTRESTO","NY"
"Neurology","SCHENECTADY","Eisai Inc.,"SHELDON","STAUNTON","6.1","Food and Beverage","Fycempa","NY"
"KINGSTON","Celgene Corporation","CHARLES","KUTLER","15.41","Food and Beverage","NY"
"RESPIRATORY","Brooklyn","Boehringer Ingelheim Pharmaceuticals, Inc.,"NABIL","RAOOF","16.36","Food and Beverage","SPIRIVA RESPIMAT","NY"
"BRONX","Allergan Inc.,"SHARON","WILLIAMS","22.14","Food and Beverage","NY"
"Gout/Rheumatology","BROOKLYN","Takeda Pharmaceuticals U.S.A., Inc.,"SIMON","RASKIN","18.71","Food and Beverage","Uloric","NY"
"VI","STATEN ISLAND","The Spectranetics Corporation","OREN","HERMAN","5.07","Food and Beverage","TURBOELITE","NY"
"HCV","BROOKLYN","Gilead Sciences Inc","ELLIOT","DEHAAN","119.45","Food and Beverage","Eplusa","NY"
"Cardiovascular & Metabolism","JACKSON HEIGHTS","Janssen Pharmaceuticals, Inc","ALVARO","OLAYO","78.33","Food and Beverage","INVOKANA","NY"
"NEUROSCIENCE:PAIN","BAY SHORE","Pfizer Inc.,"CARLA","CANAVIRE-WEBER","11.37","Food and Beverage","LYRICA","NY"
"BRONX","Gilead Sciences Inc","SUSAN","AMENECHI-ENAHORO","17.48","Food and Beverage","NY"
"Pharma Product","LYNBROOK","Smith & Nephew, Inc.,"AYESHA","MOHIUDDIN","23.56","Food and Beverage","Santyl","NY"
"VALVULOTOMY","NEW YORK","LeMaitre Vascular, Inc.,"THOMAS","BERNIK","543.6","Food and Beverage","HYDRO LEMAITRE VALVULOTOME","NY"
"Immunology","WILLIAMSVILLE","AbbVie, Inc.,"JOSEPH","NEIMAN","13.65","Food and Beverage","Humira","NY"
"Vascular","NEW YORK","Abbott Laboratories","PATRICK","LOOSER","6.95","Food and Beverage","Asahi Fielder XE coronary guide wire","NY"
"RESPIRATORY","New Hartford","Boehringer Ingelheim Pharmaceuticals, Inc.,"ADNAN","CEMER","2.69","Food and Beverage","STIOLTO","NY"
"Diabetes","FLUSHING","Novo Nordisk Inc","KENNETH","SHA","14.55","Food and Beverage","Tresiba","NY"
"HCV","BRONX","Gilead Sciences Inc","DAVID","STEIN","106.15","Food and Beverage","Harvon","NY"
"NEUROSCIENCE:PAIN","BROOKLYN","Pfizer Inc.,"STEVEN","BARKOFF","10.75","Food and Beverage","LYRICA","NY"
"WANTAGH","Merck Sharp & Dohme Corporation","JEFFREY","NOVAK","15.5","Food and Beverage","NY"

```





The image shows a screenshot of the RStudio interface. The title bar reads "RStudio" and the menu bar includes "File", "Edit", "Code", "View", "Plots", "Session", "Build", "Debug", "Profile", "Tools", and "Window". The toolbar contains icons for file operations and a search bar. The main editor window shows the following R code:

```
1 # Packt: Big Data Analytics
2 # Chapter 8 Tutorial
3
4 library(shiny)
5 library(shinydashboard)
6 library(data.table)
7 library(DT)
8 library(shinyjs)
9
10
11 cms_factor_dt <- readRDS("~/r/rulespackt/cms_factor_dt.rds")
12 cms_rules_dt <- readRDS("~/r/rulespackt/cms_rules_dt.rds")
13
14 # Define UI for application that draws a histogram
15 ui <- dashboardPage (skin="green",
16   dashboardHeader(title = "Apriori Algorithm"),
17   dashboardSidebar(
18     useShinyjs(),
19     sidebarMenu(
20       uiOutput("company"),
21       uiOutput("searchlhs"),
22       uiOutput("searchrhs"),
23       uiOutput("support2"),
24       uiOutput("confidence"),
25       uiOutput("lift"),
26       downloadButton('downloadMatchingRules', "Download Rules")
```

Name	Size	Kind
app.R	6 KB	Rez Source
cms_factor_dt.rds	5.1 MB	R Data File
cms_rules_dt.rds	218 KB	R Data File
cms_rules.rds	288 KB	R Data File
www	--	Folder
fonts	--	Folder
rubik	--	Folder
fonts.css	2 KB	Casca...cument
packt2.css	6 KB	Casca...cument
webfonts	--	Folder
34D579_0_0.eot	288 KB	Document
34D579_0_0.svg	2.1 MB	SVG Image
34D579_0_0.ttf	814 KB	TrueType® font
34D579_0_0.woff	386 KB	Document
34D579_0_0.woff2	265 KB	Document
svg_test.html	2 KB	HTML

The screenshot shows the RStudio interface. The code editor contains the following R code:

```

1 # Packt: Big Data Analytics
2 # Chapter 8 Tutorial
3
4 library(shiny)
5 library(shinydashboard)
6 library(data.table)
7 library(DT)
8 library(shinyjs)
9
10
11 cms_factor_dt <- readRDS("~/rulespackt/cms_factor_dt.rds")
12 cms_rules_dt <- readRDS("~/rulespackt/cms_rules_dt.rds")
13
14 # Define UI for application that draws a histogram
15 ui <- dashboardPage (skin="green",

```

The 'Run App' dropdown menu is open, showing the following options:

- Run in Window (checked)
- Run in RStudio
- Run External

Apriori Algorithm

Select Company

Select All

Search LHS

Search

Search RHS

Search

Support

0.01

0 0.005 0.01 0.02 0.03 0.04

Confidence

0.5

0 0.1 0.2 0.4 0.6 0.8 1

Lift

1

0 1 2 3 4 5 6 7 8 9 10

[Download Rules](#)

Show 10 entries

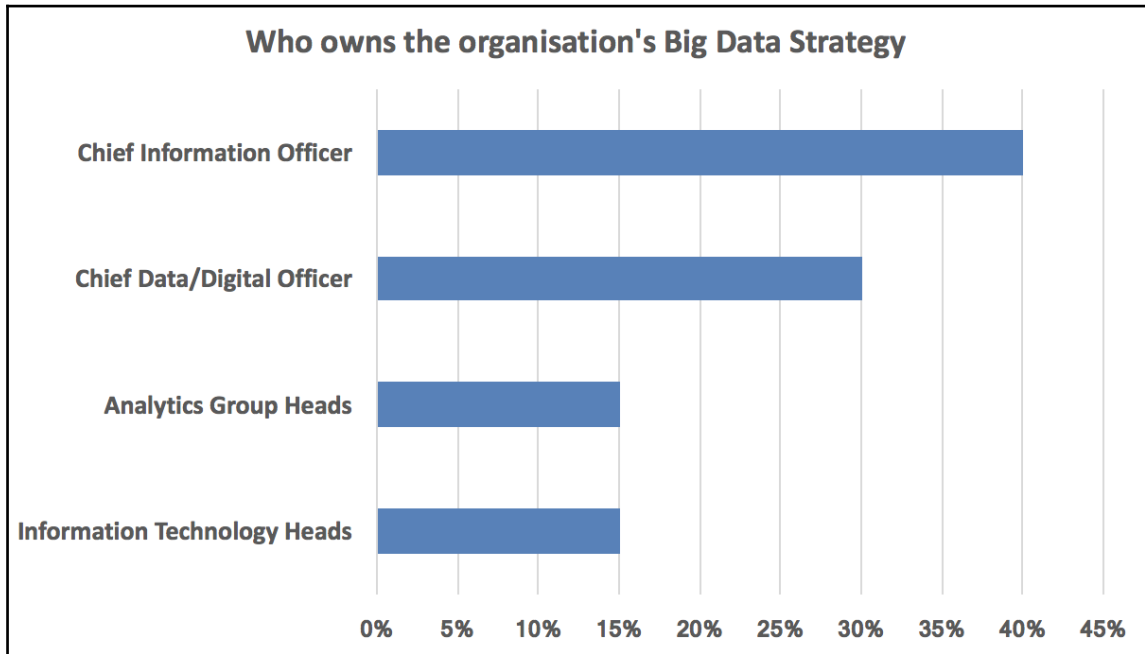
Search:

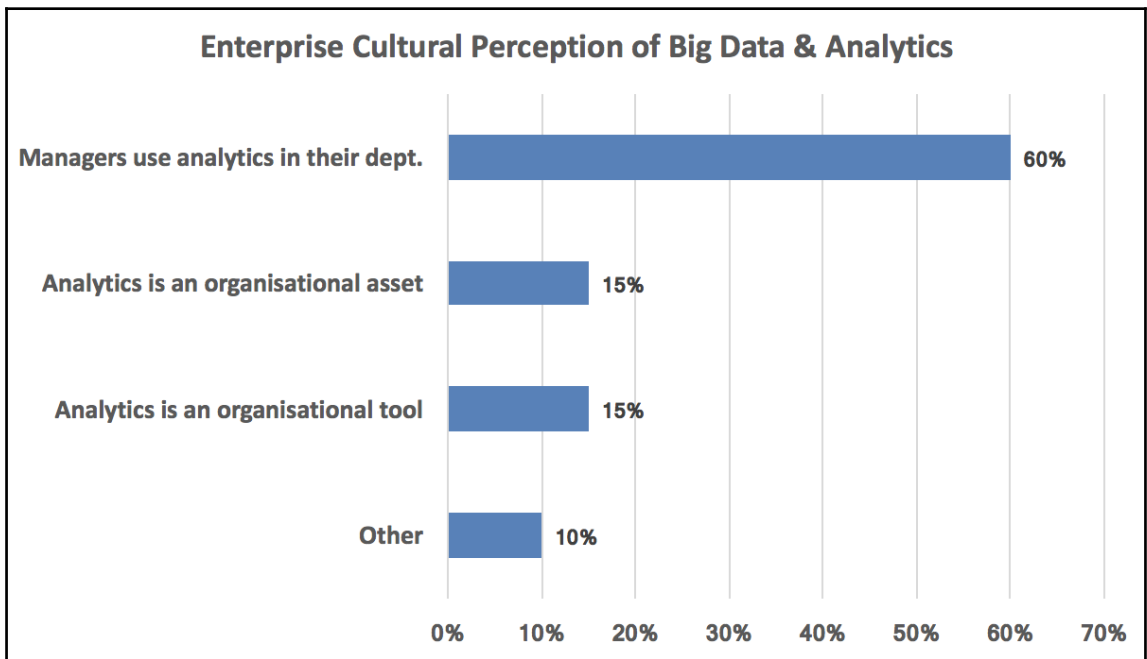
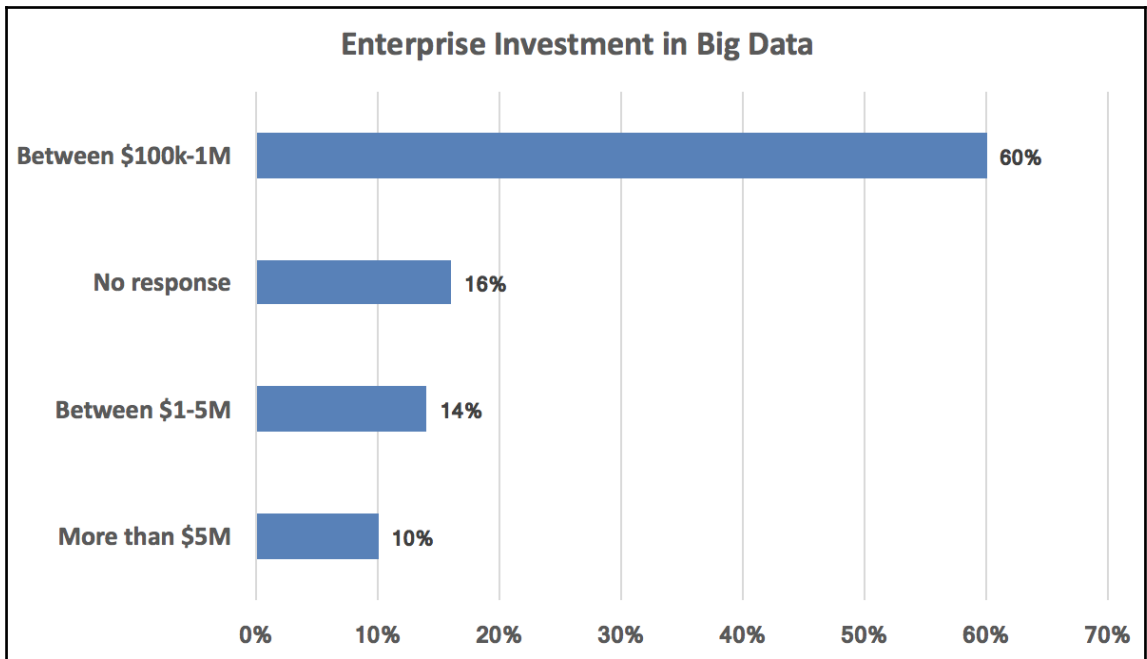
	LHS	RHS	Support	Confidence	Lift
1	{company=SANOFI AND GENZYME US COMPANIES,product=TOUJEO}	{category=DIABETES}	0.02	1	14.38
2	{category=DIABETES,product=TOUJEO}	{company=SANOFI AND GENZYME US COMPANIES}	0.02	1	35.45
3	{category=DIABETES,company=SANOFI AND GENZYME US COMPANIES}	{product=TOUJEO}	0.02	0.9	60.2
4	{category=CARDIOVASCULAR & METABOLISM,product=INVOKANA}	{company=JANSSEN PHARMACEUTICALS, INC}	0.02	1	16.93
5	{company=JANSSEN PHARMACEUTICALS, INC,product=INVOKANA}	{category=CARDIOVASCULAR & METABOLISM}	0.02	1	24.43
6	{category=CARDIOVASCULAR AND METABOLISM,product=FARXIGA}	{company=ASTRAZENECA PHARMACEUTICALS LP}	0.02	1	19.1
7	{company=ASTRAZENECA PHARMACEUTICALS LP,product=FARXIGA}	{category=CARDIOVASCULAR AND METABOLISM}	0.02	0.99	32.15
8	{category=CARDIOVASCULAR AND METABOLISM,company=ASTRAZENECA PHARMACEUTICALS LP}	{product=FARXIGA}	0.02	0.58	32.15
9	{category=CARDIOVASCULAR AND METABOLISM,product=FARXIGA}	{paymentNature=FOOD AND BEVERAGE}	0.02	0.96	1.1
10	{paymentNature=FOOD AND BEVERAGE,product=FARXIGA}	{category=CARDIOVASCULAR AND METABOLISM}	0.02	0.99	32.13

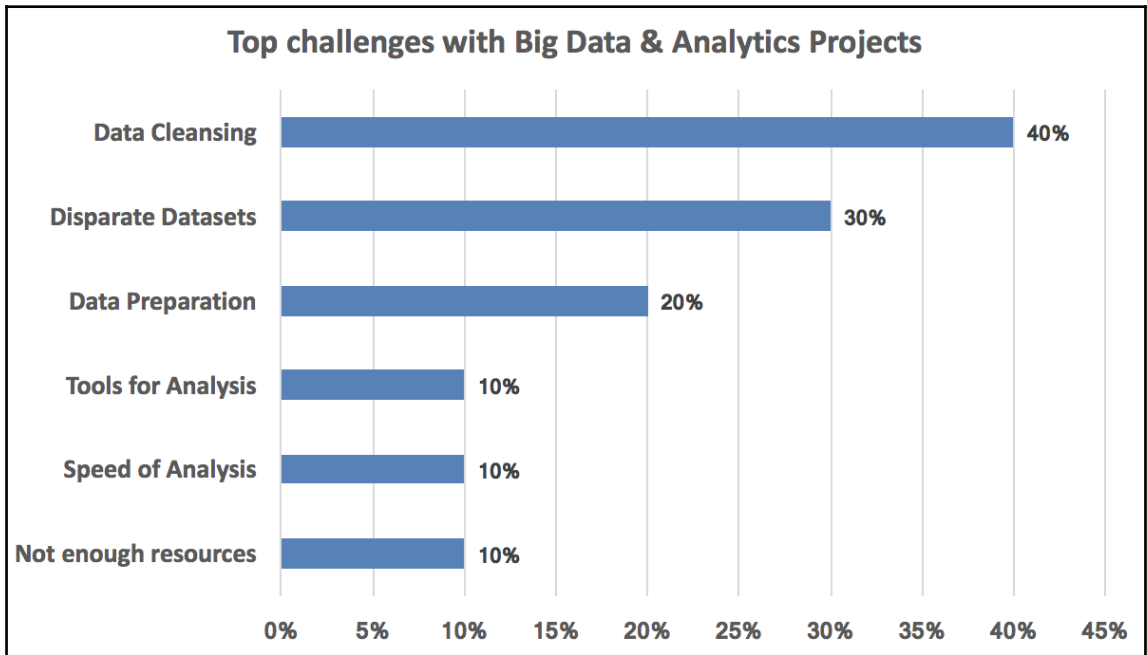
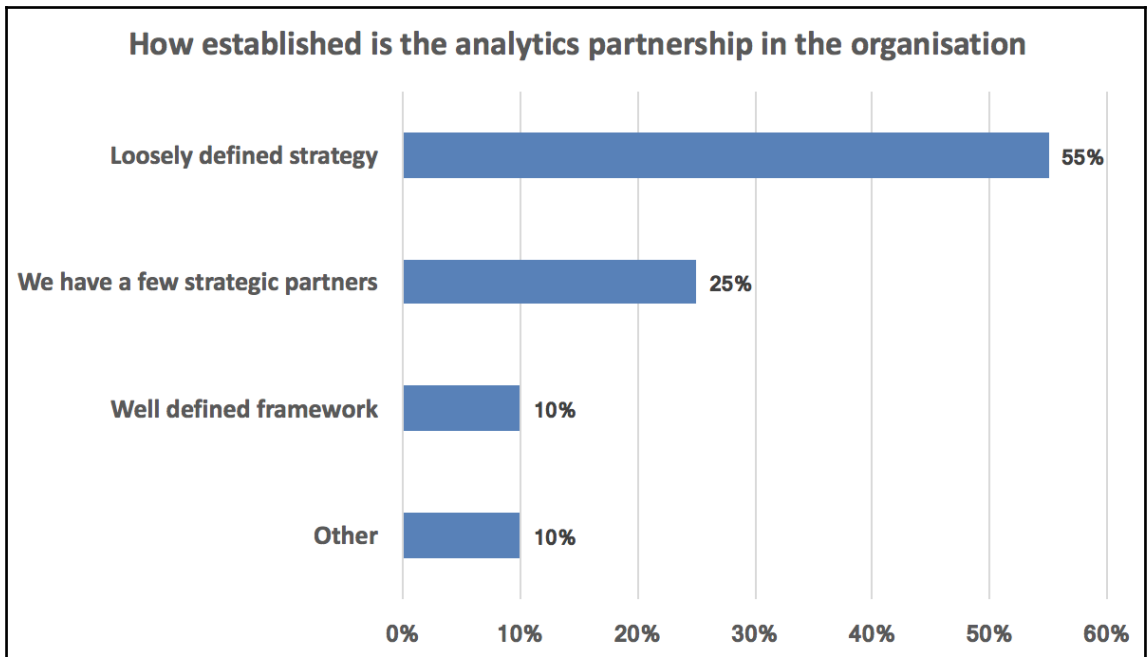
Showing 1 to 10 of 77 entries

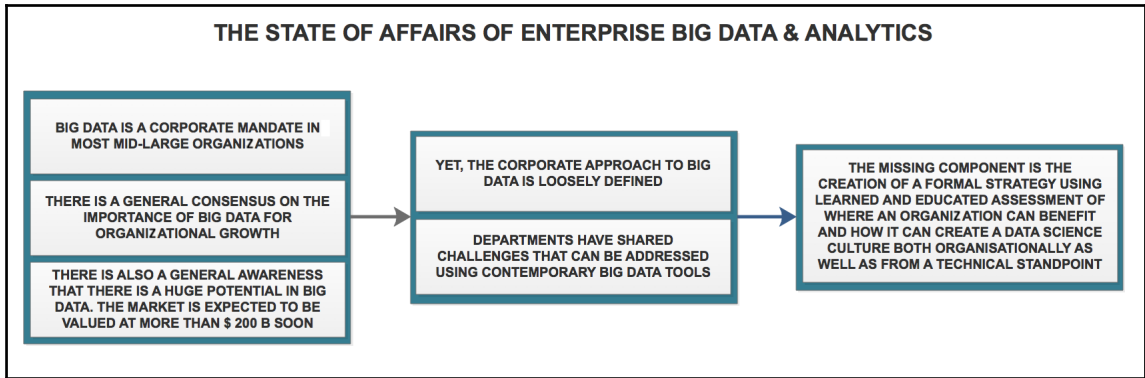
Previous 1 2 3 4 5 - 8 Next

Chapter 9: Enterprise Data Science










Amazon Redshift Spectrum

Amazon Redshift Spectrum enables you to run Amazon Redshift SQL queries against exabytes of data in Amazon S3. With Redshift Spectrum, you can extend the analytic power of Amazon Redshift beyond data stored on local disks in your data warehouse to query vast amounts of unstructured data in your Amazon S3 "data lake" -- without having to load or transform any data. Redshift Spectrum applies sophisticated query optimization, scaling processing across thousands of nodes so results are fast -- even with large data sets and complex queries.

Redshift Spectrum directly queries data in Amazon S3 using the open data formats you already use, including Avro, CSV, Grok, ORC, Parquet, RCFile, RegexSerDe, SequenceFile, TextFile, and TSV. Since Redshift Spectrum supports the same SQL syntax of Amazon Redshift, you can run sophisticated queries using the same Business Intelligence (BI) tools you use today. You can also run queries that span both the frequently accessed data stored locally in Amazon Redshift and your full data sets stored cost-effectively in Amazon S3.

Redshift Spectrum gives you the freedom to store your data where you want, in the format you want, and have it available for processing when you need it. There are no up-front payments or commitments with Redshift Spectrum; you only pay for the queries you run.

[Request support for your proof-of-concept or evaluation »](#)


?
N

COMPOSE QUERY

Query History

Job History

Please create a dataset or select a new project from the menu above.

- ▼ bigquery-public-data
 - ▶ baseball
 - ▶ bls
 - ▶ census_bureau_international
 - ▶ census_bureau_usa
 - ▶ chicago_crime
 - ▶ chicago_taxi_trips
 - ▶ cloud_storage_geo_index
 - ▶ cms_codes
 - ▶ cms_medicare
 - home_health_agencies_2013
 - home_health_agencies_2014
 - hospice_providers_2014
 - hospital_general_info
 - inpatient_charges_2011
 - inpatient_charges_2012
 - inpatient_charges_2013
 - inpatient_charges_2014
 - inpatient_charges_2015
 - nursing_facilities_2013
 - nursing_facilities_2014
 - outpatient_charges_2011

New Query Query Editor UDF Editor X

```
1 SELECT count(*) FROM [bigquery-public-data:world_bank_health_population.country_summary];
```

SQL

RUN QUERY
Save Query
Save View
Format Query
Show Options

Ctrl + Enter: run query, Tab or Ctrl + Space: autocomplete. ✔

Table Details: outpatient_charges_2015 Refresh Query Table Copy Table Export Table Delete Table

Schema
Details
Preview

Row	provider_id	provider_name	provider_street_address	provider_city	provider_state	provider_zipcode	apc
1	010001	Southeast Alabama Medical Center	1108 Ross Clark Circle	Dothan	AL	36301	0012 - Level I Debrider
2	010005	Marshall Medical Center South	2505 U S Highway 431 North	Boaz	AL	35957	0012 - Level I Debrider
3	010007	Mizell Memorial Hospital	702 N Main St	Opp	AL	36467	0012 - Level I Debrider
4	010008	Crenshaw Community Hospital	101 Hospital Circle	Luverne	AL	36049	0012 - Level I Debrider
5	010016	Shelby Baptist Medical Center	1000 First Street North	Alabaster	AL	35007	0012 - Level I Debrider
6	010024	Jackson Hospital & Clinic Inc	1725 Pine Street	Montgomery	AL	36106	0012 - Level I Debrider
7	010033	University Of Alabama Hospital	619 South 19th Street	Birmingham	AL	35233	0012 - Level I Debrider
8	010035	Cullman Regional Medical Center	1912 Alabama Highway 157	Cullman	AL	35058	0012 - Level I Debrider
9	010038	Stringfellow Memorial Hospital	301 East 18th St	Anniston	AL	36201	0012 - Level I Debrider
10	010040	Gadsden Regional Medical Center	1007 Goodyear Avenue	Gadsden	AL	35903	0012 - Level I Debrider
11	010046	Riverview Regional Medical Center	600 South Third Street	Gadsden	AL	35901	0012 - Level I Debrider
12	010055	Flowers Hospital	4370 West Main Street	Dothan	AL	36305	0012 - Level I Debrider

Table JSON
First < Prev
Rows 1 - 12 of 32532
Next > Last

[127]

The screenshot shows the Microsoft Azure website's 'Azure solutions' page. At the top, there is a navigation bar with the Microsoft Azure logo, a phone number (SALES 1-800-867-1389), and links for 'MY ACCOUNT', 'PORTAL', and a search bar. Below the navigation bar, the main heading is 'Azure solutions'. A sub-heading reads: 'We've grouped Azure services, third-party applications, and related products together to help meet the most common business needs and scenarios—including yours. Access architectures, tutorials, documentation, examples, templates, partners, and other resources to get started quickly.' A green 'Start free >' button is positioned below the text. To the right, a dark box contains the text: 'Design and implement secure, highly-available solutions' and a link 'View solution architectures >'. Below this is a row of partner logos: NBC, GE Healthcare, 3M, DocuSign, BMW, AccuWeather, and Heineken. The main content area is divided into eight categories, each with an icon, a title, and a brief description:

- Digital marketing**: Connect with customers worldwide with digital campaigns that are personalized and scalable.
- Mobile**: Reach your customers everywhere, on every device, with a single mobile app build.
- E-commerce**: Give customers what they want with a personalized, scalable, and secure shopping experience.
- LOB applications**: Modernize your internal line of business (LOB) apps to meet today's IT challenges.
- SharePoint on Azure**: Deploy SharePoint servers rapidly and scale as needed with a cost-effective infrastructure.
- Dynamics on Azure**: Fuel business growth by bringing together enterprise resource planning (ERP) and cloud services.
- SAP on Azure**: Bring cloud scale and agility to your mission-critical SAP workloads.
- Red Hat on Azure**: Achieve hybrid cloud agility for your enterprise with Red Hat solutions on Azure.

scikit-learn.org/stable/

Home Installation Documentation Examples

Google Custom Search Search

Fork me on GitHub

scikit-learn

Machine Learning in Python

- Simple and efficient tools for data mining and data analysis
- Accessible to everybody, and reusable in various contexts
- Built on NumPy, SciPy, and matplotlib
- Open source, commercially usable - BSD license

Classification

Identifying to which category an object belongs to.

Applications: Spam detection, Image recognition.

Algorithms: SVM, nearest neighbors, random forest, ... — Examples

Regression

Predicting a continuous-valued attribute associated with an object.

Applications: Drug response, Stock prices.

Algorithms: SVR, ridge regression, Lasso, ... — Examples

Clustering

Automatic grouping of similar objects into sets.

Applications: Customer segmentation, Grouping experiment outcomes

Algorithms: k-Means, spectral clustering, mean-shift, ... — Examples

Dimensionality reduction

Reducing the number of random variables to consider.

Applications: Visualization, Increased efficiency

Algorithms: PCA, feature selection, non-negative matrix factorization. — Examples

Model selection

Comparing, validating and choosing parameters and models.

Goal: Improved accuracy via parameter tuning

Modules: grid search, cross validation, metrics. — Examples

Preprocessing

Feature extraction and normalization.

Application: Transforming input data such as text for use with machine learning algorithms.

Modules: preprocessing, feature extraction. — Examples

News

On-going development: What's new (Changelog)

October 2017. scikit-learn 0.19.1 is available for download (Changelog).

Community


About us See authors and contributing

More Machine Learning Find related projects

Questions? See FAQ and stackoverflow

Who uses scikit-learn?

EVERNOTE

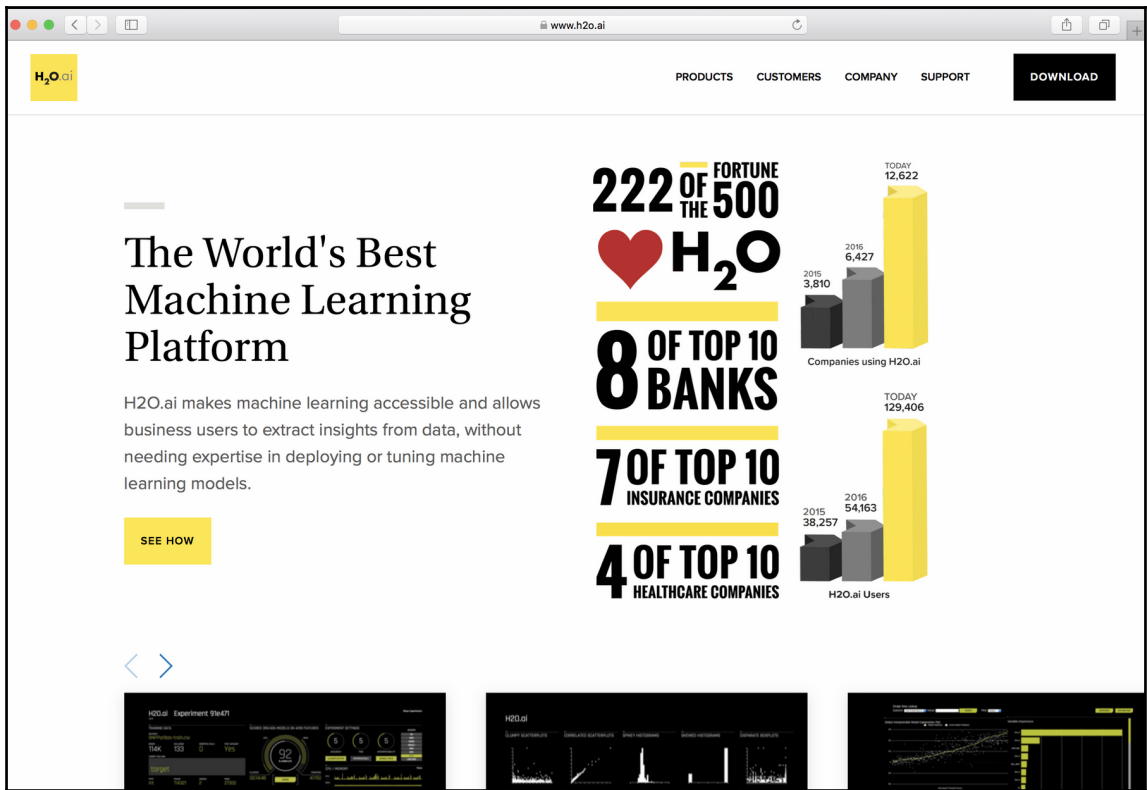
 [Overview](#) [Programming Guides](#) [API Docs](#) [Deploying](#) [More](#)

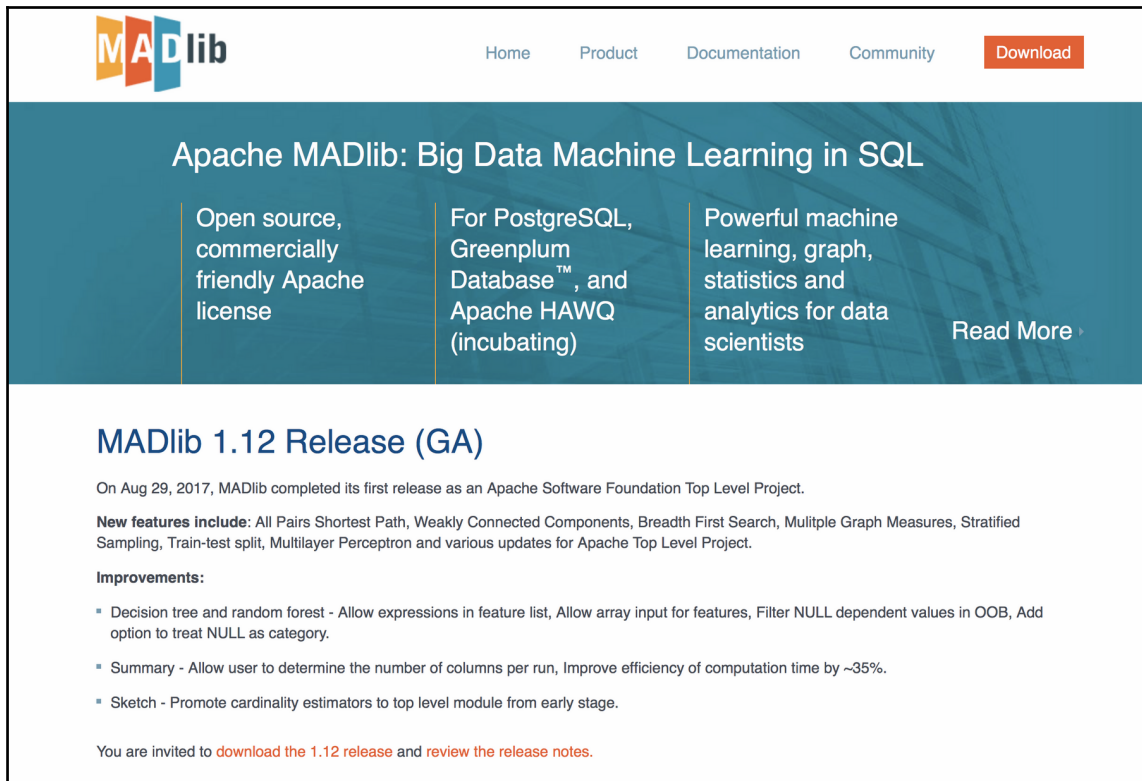
Machine Learning Library (MLlib)

MLlib is Spark's scalable machine learning library consisting of common learning algorithms and utilities, including classification, regression, clustering, collaborative filtering, dimensionality reduction, as well as underlying optimization primitives, as outlined below:

- [Data types](#)
- [Basic statistics](#)
 - summary statistics
 - correlations
 - stratified sampling
 - hypothesis testing
 - random data generation
- [Classification and regression](#)
 - linear models (SVMs, logistic regression, linear regression)
 - decision trees
 - naive Bayes
- [Collaborative filtering](#)
 - alternating least squares (ALS)
- [Clustering](#)
 - k-means
- [Dimensionality reduction](#)
 - singular value decomposition (SVD)
 - principal component analysis (PCA)
- [Feature extraction and transformation](#)
- [Optimization \(developer\)](#)
 - stochastic gradient descent
 - limited-memory BFGS (L-BFGS)

MLlib is under active development. The APIs marked `ExperimentalDeveloperApi` may change in future releases, and the migration guide below will explain all changes between releases.





The screenshot shows the Apache MADlib website. At the top left is the MADlib logo. To its right are navigation links: Home, Product, Documentation, and Community. A red 'Download' button is on the far right. Below the navigation is a dark teal banner with the title 'Apache MADlib: Big Data Machine Learning in SQL'. Underneath the banner are three columns of text: 'Open source, commercially friendly Apache license', 'For PostgreSQL, Greenplum Database™, and Apache HAWQ (incubating)', and 'Powerful machine learning, graph, statistics and analytics for data scientists'. A 'Read More' link with a right-pointing arrow is on the right side of the banner. Below the banner is a white section titled 'MADlib 1.12 Release (GA)'. The text below the title states: 'On Aug 29, 2017, MADlib completed its first release as an Apache Software Foundation Top Level Project.' It then lists 'New features include: All Pairs Shortest Path, Weakly Connected Components, Breadth First Search, Multiple Graph Measures, Stratified Sampling, Train-test split, Multilayer Perceptron and various updates for Apache Top Level Project.' Under 'Improvements:', there are three bullet points: 'Decision tree and random forest - Allow expressions in feature list, Allow array input for features, Filter NULL dependent values in OOB, Add option to treat NULL as category.', 'Summary - Allow user to determine the number of columns per run, Improve efficiency of computation time by ~35%', and 'Sketch - Promote cardinality estimators to top level module from early stage.' At the bottom of this section, it says 'You are invited to [download the 1.12 release](#) and [review the release notes](#).'

IBM Search [User Icon] [Menu Icon]

Cloud Why IBM? Products Solutions Garage Pricing Blogs Docs Support [Sign up](#)

Unlock the power of cognitive with IBM Watson

Turn unstructured data into intelligence and competitive advantages

[Take your first step](#) [View all Watson offerings](#)

Oh, the things you can do with Watson on the IBM Cloud

- Tap into unstructured data**
IBM Watson® processes key insights, patterns and relationships across unstructured images, emails, social media and more.
- Improve decisions based on real-time trends**
Reveal valuable insights from social trends in real time with the solutions-integrating IBM Watson Natural Language Classifier Service.
- Quickly and securely build truly cognitive applications**
IBM Watson Conversation Service lets you speed time to market and create better customer experiences by understanding and responding to customers.

Talk to an expert




The screenshot shows a web browser window at www.openstack.org/software/. The main heading is "What is OpenStack?". Below it is a diagram illustrating the OpenStack architecture. At the top, "YOUR APPLICATIONS" (represented by three "APP" boxes) connect to the "OPENSTACK Control Plane" via "APIs". The Control Plane manages various resources: "Bare Metal", "Virtual Machines", "Containers", "Object Storage", "File Storage", and "Block Storage". These resources are connected to a "COMMON NETWORK". On the right side, there are two additional API-related components: "Dashboard (GUI)" and "Monitoring & Tools".

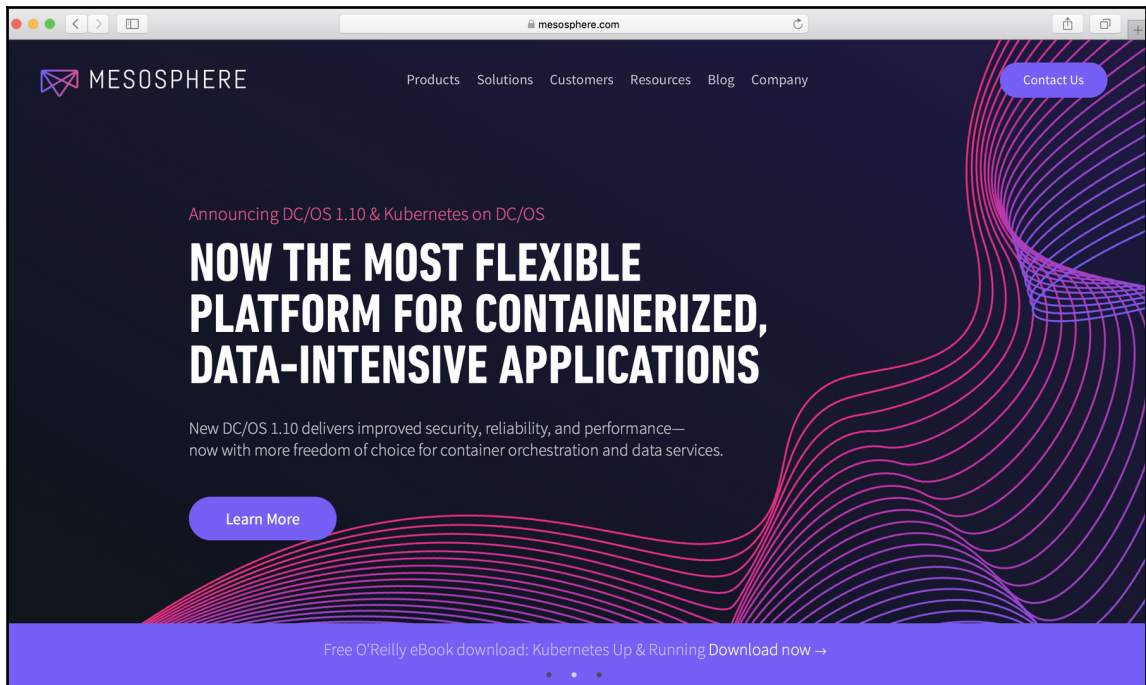
OpenStack is a cloud operating system that controls large pools of compute, storage, and networking resources throughout a datacenter, all managed through a dashboard that gives administrators control while empowering their users to provision resources through a web interface.

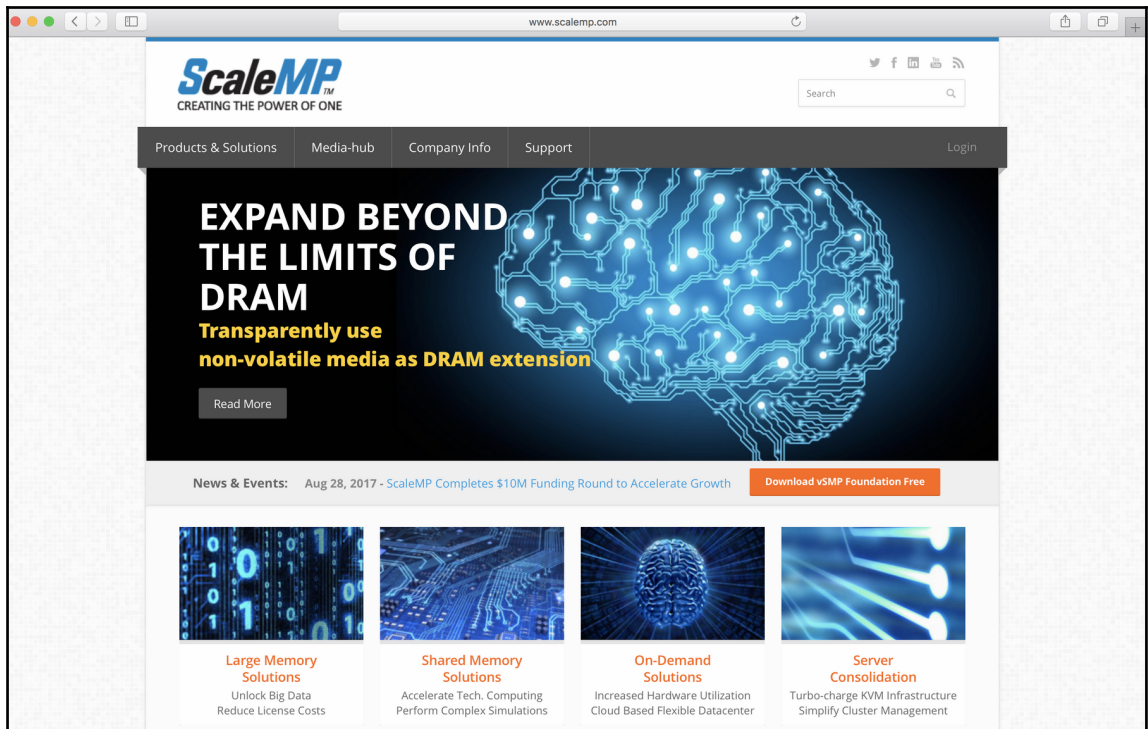
Learn more about OpenStack's compute, storage, networking, take a tour of the dashboard, or dive in and learn how to get started with OpenStack today.

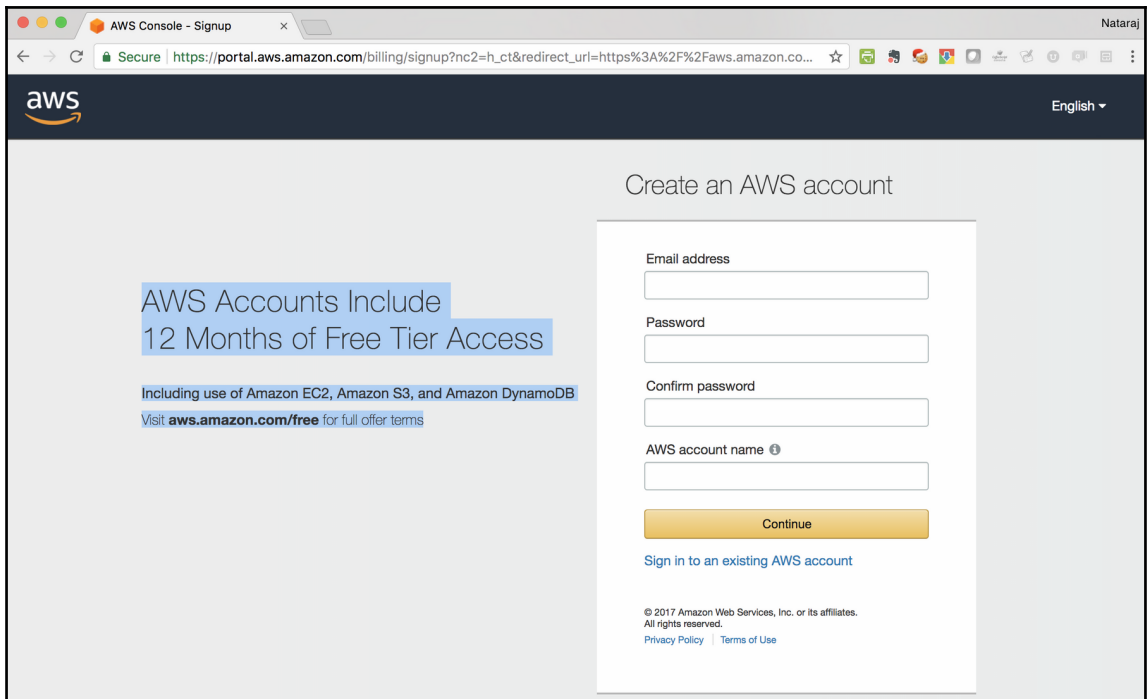
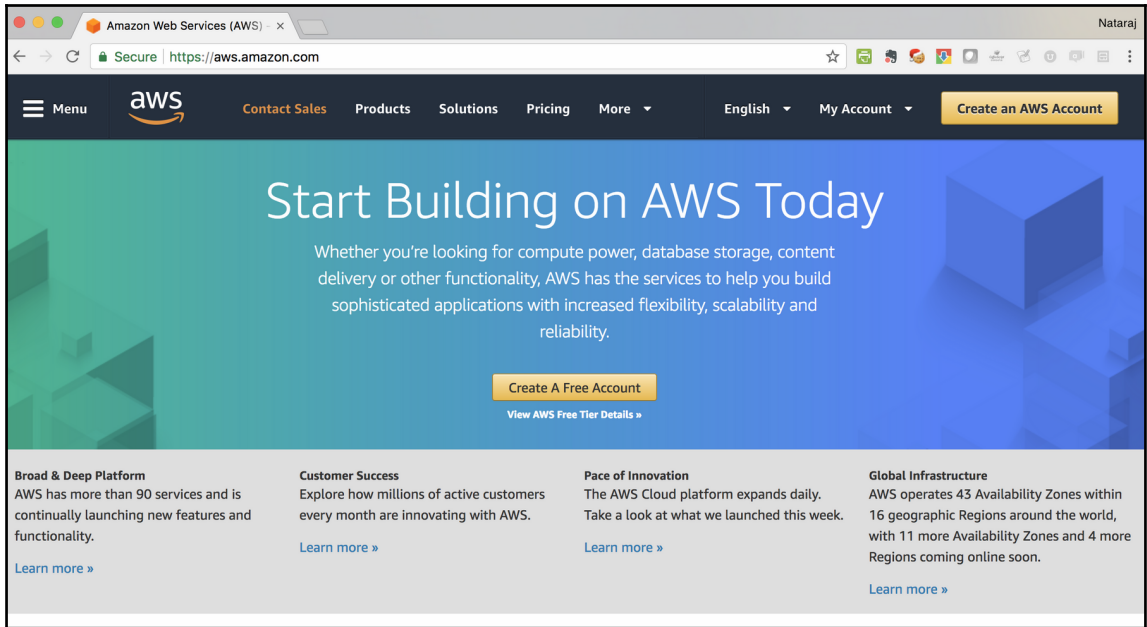
If you're looking for OpenStack related products and services such as Distributions, Appliances, Public Clouds, Consultants, or Training check out our OpenStack Marketplace.

What can OpenStack Automate?

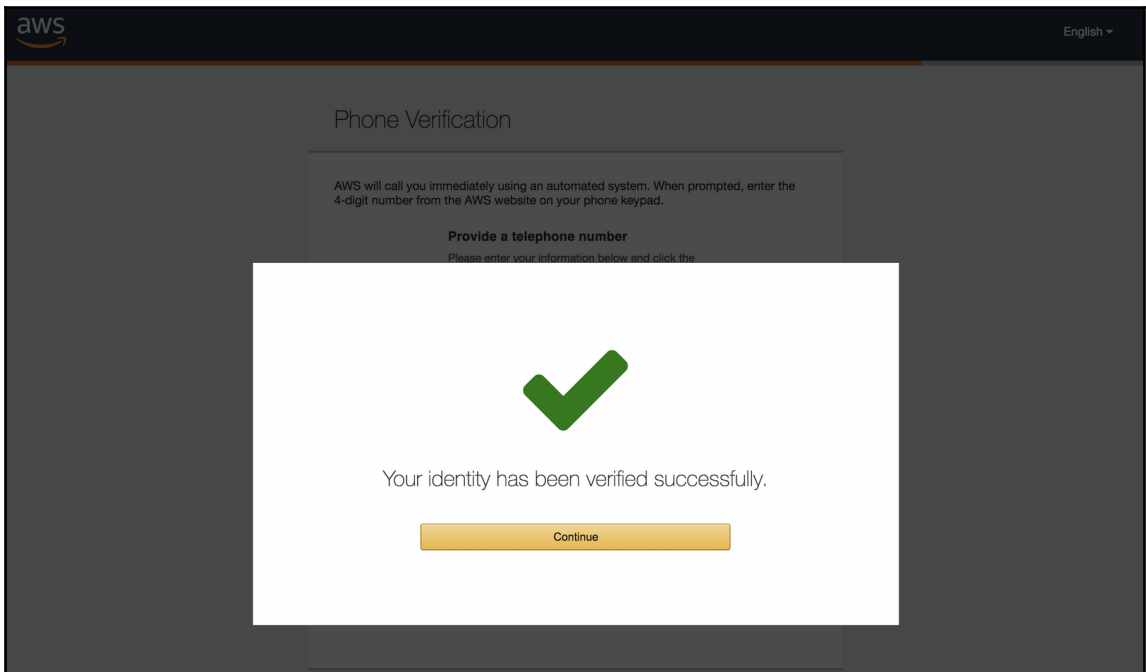
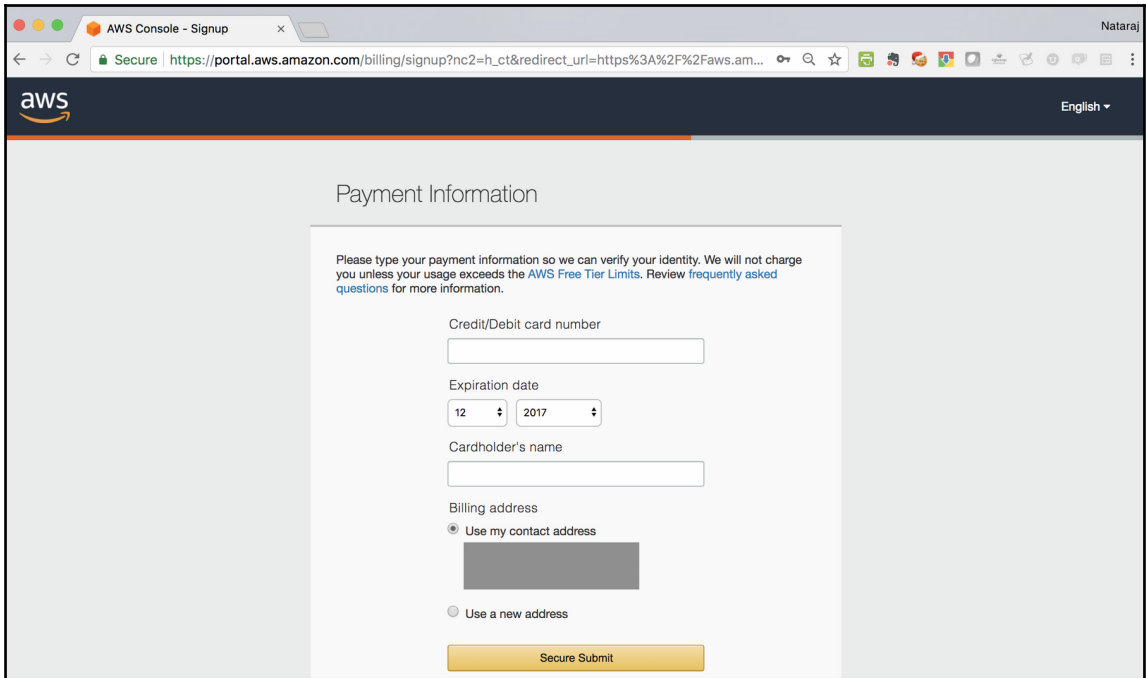
 Compute 6 OpenStack Projects	 Storage, Backup & Recovery 5 OpenStack Projects	 Networking & Content Delivery 7 OpenStack Projects
--	---	--







The screenshot shows a web browser window with the title 'AWS Console - Signup'. The address bar shows a secure connection to 'https://portal.aws.amazon.com/billing/signup?nc2=h_ct&redirect_url=https%3A%2F%2Faws.am...'. The AWS logo is in the top left, and 'English' is in the top right. The main content area is titled 'Contact Information' with a note 'All fields are required.' Below this is a form with the instruction 'Please select the account type and complete the fields below with your contact details.' The form includes: 'Account type' with radio buttons for 'Professional' and 'Personal' (selected); 'Full name' with a text box containing 'Nataraj Dasgupta'; 'Phone number' with a text box containing 'Enter Phone Number Here'; 'Country' with a dropdown menu set to 'United States'; 'Address' with a text box containing 'Enter Address Here' and a smaller box below it with the placeholder 'Apartment, suite, unit, building, floor, etc.'; 'City' with a text box; 'State / Province or region' with a text box; 'Postal code' with a text box; a checkbox for 'Check here to indicate that you have read and agree to the terms of the AWS Customer Agreement'; and a yellow 'Create Account and Continue' button at the bottom.



Select a Support Plan

AWS offers a selection of support plans to meet your needs. Choose the support plan that best aligns with your AWS usage. [Learn more](#)

Basic Plan	Developer Plan	Business Plan
Free	From \$29/month	From \$100/month
<ul style="list-style-type: none">Included with all accounts24/7 self-service access to forums and resourcesBest practice checks to help improve security and performanceAccess to health status and notifications	<ul style="list-style-type: none">For early adoption, testing and developmentEmail access to AWS Support during business hours1 primary contact can open an unlimited number of support cases12-hour response time for nonproduction systems	<ul style="list-style-type: none">For production workloads & business-critical dependencies24/7 chat, phone, and email access to AWS SupportUnlimited contacts can open an unlimited number of support cases1-hour response time for production systems

Need Enterprise level support?
Contact your account manager for additional information on running business and mission critical-workloads on AWS (starting at \$15,000/month). [Learn more](#)

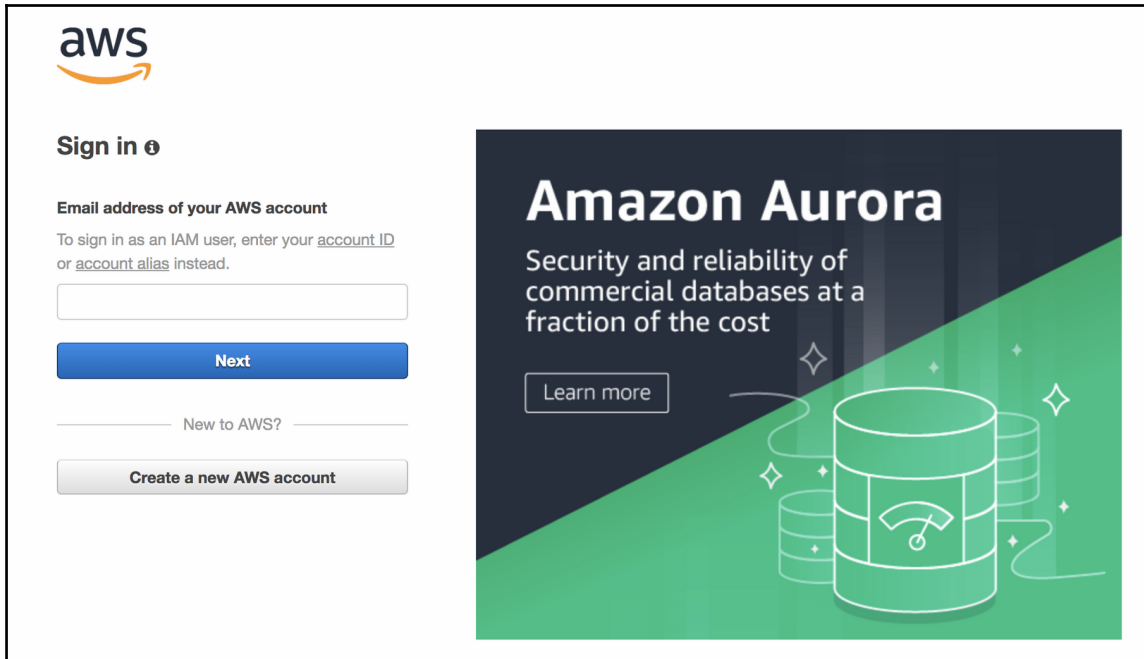
Welcome to Amazon Web Services

Thank you for creating an Amazon Web Services Account. We are activating your account, which should only take a few minutes. You will receive an email when this is complete.

[Sign in to the Console](#)

[Contact Sales](#)

© 2017 Amazon Web Services, Inc. or its affiliates. All rights reserved.
[Privacy Policy](#) | [Terms of Use](#) | [Sign Out](#)



aws

Sign in

Email address of your AWS account

To sign in as an IAM user, enter your [account ID](#) or [account alias](#) instead.

Next

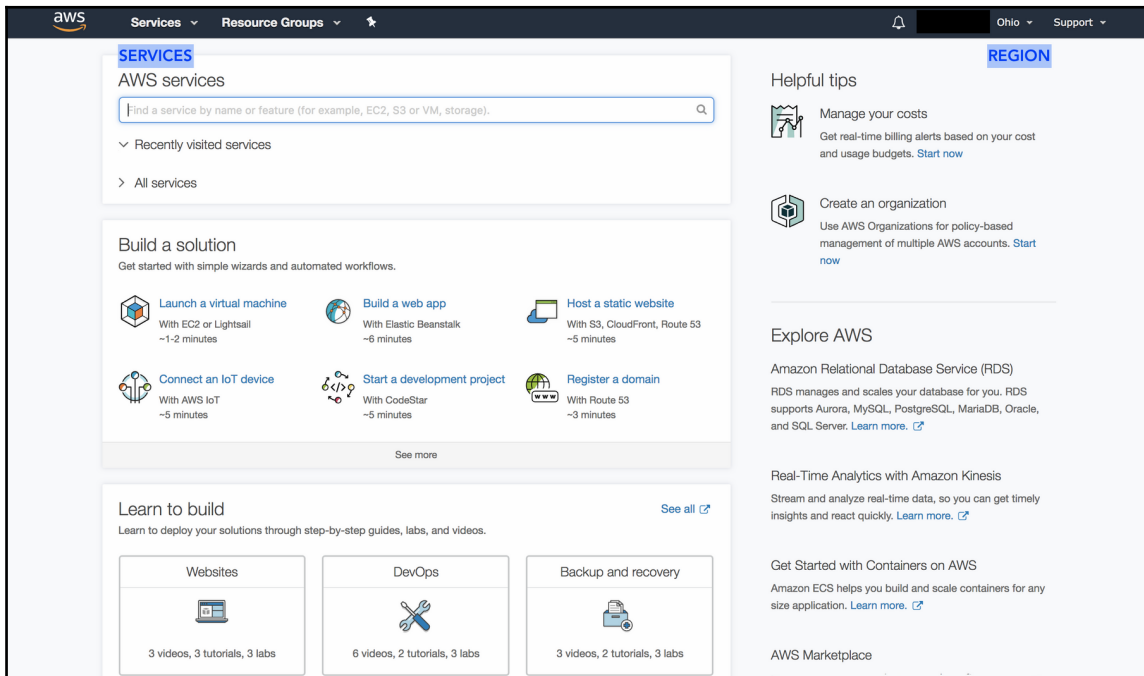
New to AWS?

Create a new AWS account

Amazon Aurora

Security and reliability of commercial databases at a fraction of the cost

[Learn more](#)



aws Services Resource Groups Ohio Support

SERVICES

AWS services

Find a service by name or feature (for example, EC2, S3 or VM, storage).

Recently visited services

All services

Build a solution

Get started with simple wizards and automated workflows.

- Launch a virtual machine**
With EC2 or Lightsail
~1-2 minutes
- Build a web app**
With Elastic Beanstalk
~6 minutes
- Host a static website**
With S3, CloudFront, Route 53
~5 minutes
- Connect an IoT device**
With AWS IoT
~5 minutes
- Start a development project**
With CodeStar
~5 minutes
- Register a domain**
With Route 53
~3 minutes

See more

Learn to build

Learn to deploy your solutions through step-by-step guides, labs, and videos. [See all](#)

- Websites**
3 videos, 3 tutorials, 3 labs
- DevOps**
6 videos, 2 tutorials, 3 labs
- Backup and recovery**
3 videos, 2 tutorials, 3 labs

Helpful tips

Manage your costs
Get real-time billing alerts based on your cost and usage budgets. [Start now](#)

Create an organization
Use AWS Organizations for policy-based management of multiple AWS accounts. [Start now](#)

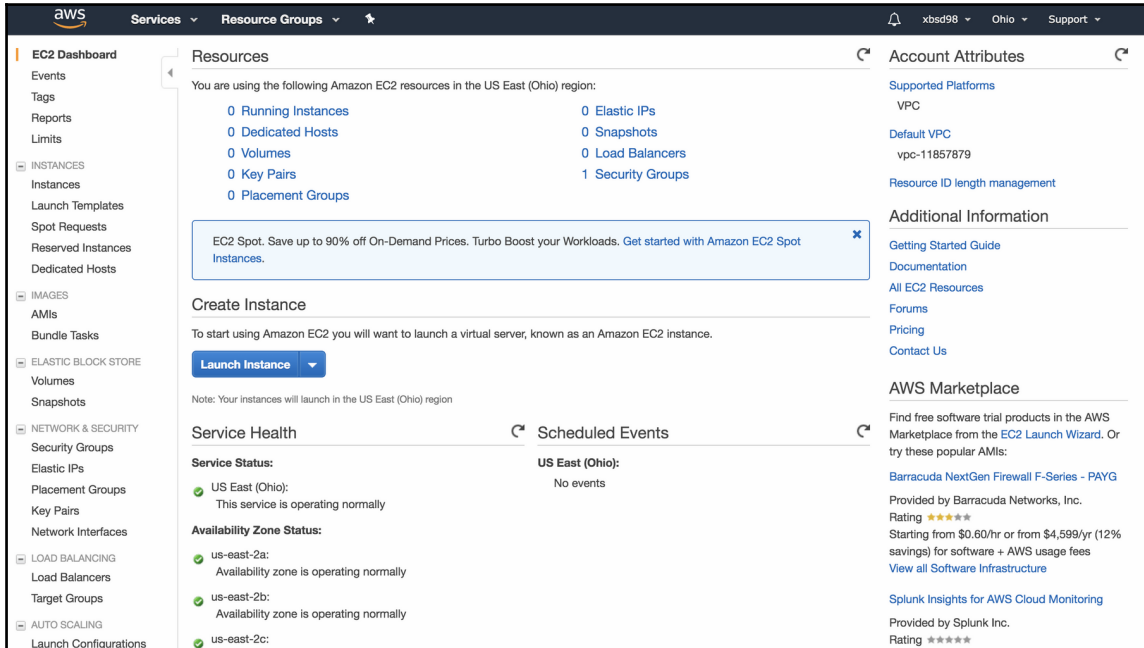
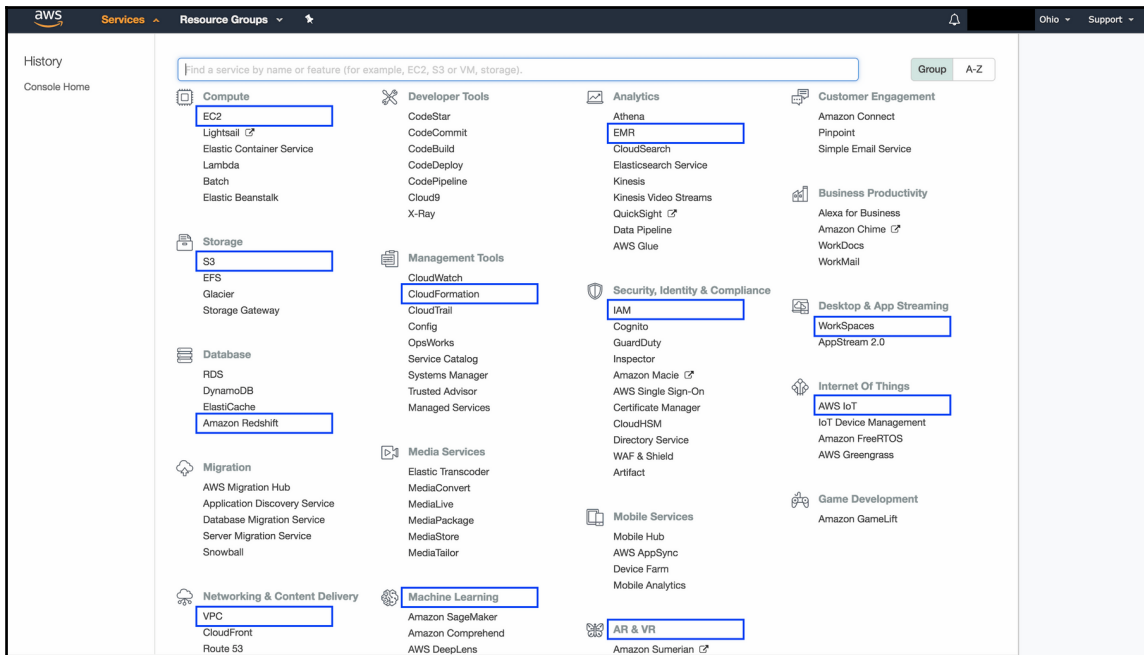
Explore AWS

Amazon Relational Database Service (RDS)
RDS manages and scales your database for you. RDS supports Aurora, MySQL, PostgreSQL, MariaDB, Oracle, and SQL Server. [Learn more](#)

Real-Time Analytics with Amazon Kinesis
Stream and analyze real-time data, so you can get timely insights and react quickly. [Learn more](#)

Get Started with Containers on AWS
Amazon ECS helps you build and scale containers for any size application. [Learn more](#)

AWS Marketplace



Step 1: Choose an Amazon Machine Image (AMI) Cancel and Exit

An AMI is a template that contains the software configuration (operating system, application server, and applications) required to launch your instance. You can select an AMI provided by AWS, our user community, or the AWS Marketplace; or you can select one of your own AMIs.

1. Choose AMI 2. Choose Instance Type 3. Configure Instance 4. Add Storage 5. Add Tags 6. Configure Security Group 7. Review

Quick Start 1 to 34 of 34 AMIs

- My AMIs
- AWS Marketplace
- Community AMIs

Free tier only

Amazon Linux Free tier eligible	Amazon Linux AMI 2017.09.1 (HVM), SSD Volume Type - ami-15e9c770 The Amazon Linux AMI is an EBS-backed, AWS-supported image. The default image includes AWS command line tools, Python, Ruby, Perl, and Java. The repositories include Docker, PHP, MySQL, PostgreSQL, and other packages. Root device type: ebs Virtualization type: hvm	Select 64-bit
Red Hat Free tier eligible	Red Hat Enterprise Linux 7.4 (HVM), SSD Volume Type - ami-cfdafaaa Red Hat Enterprise Linux version 7.4 (HVM), EBS General Purpose (SSD) Volume Type Root device type: ebs Virtualization type: hvm	Select 64-bit
SUSE Linux Free tier eligible	SUSE Linux Enterprise Server 12 SP3 (HVM), SSD Volume Type - ami-36f5db53 SUSE Linux Enterprise Server 12 Service Pack 3 (HVM), EBS General Purpose (SSD) Volume Type. Public Cloud, Advanced Systems Management, Web and Scripting, and Legacy modules enabled. Root device type: ebs Virtualization type: hvm	Select 64-bit
Ubuntu Free tier eligible	Ubuntu Server 16.04 LTS (HVM), SSD Volume Type - ami-82f4dae7 Ubuntu Server 16.04 LTS (HVM),EBS General Purpose (SSD) Volume Type. Support available from Canonical (http://www.ubuntu.com/cloud/services). Root device type: ebs Virtualization type: hvm	Select 64-bit
Windows Free tier eligible	Microsoft Windows Server 2016 Base - ami-21587144 Microsoft Windows 2016 Datacenter edition. [English] Root device type: ebs Virtualization type: hvm	Select 64-bit

Step 1: Choose an Amazon Machine Image (AMI) Cancel and Exit

An AMI is a template that contains the software configuration (operating system, application server, and applications) required to launch your instance. You can select an AMI provided by AWS, our user community, or the AWS Marketplace; or you can select one of your own AMIs.

1. Choose AMI 2. Choose Instance Type 3. Configure Instance 4. Add Storage 5. Add Tags 6. Configure Security Group 7. Review

Quick Start 1 to 6 of 6 AMIs

Search:

Community AMIs

- Operating system
 - Amazon Linux
 - Cent OS
 - Debian
 - Fedora
 - Gentoo
 - OpenSUSE
 - Other Linux
 - Red Hat
 - SUSE Linux
 - Ubuntu
 - Windows
- Architecture
 - 32-bit
 - 64-bit
- Root device type
 - EBS

awsmarketplace	7 results for "RStudio" on AWS Marketplace Partner software pre-configured to run on AWS	
Community AMIs	RStudio-1.0.153_R-3.4.1_Julia-0.6.0_ubuntu-16.04-LTS-64bit - ami-12d1f277 Ready to run RStudio + (experimental) Julia server for statistical computation (www.louisaslett.com). Connect to instance public DNS in web browser (standard port 80), username rstudio and password rstudio Root device type: ebs Virtualization type: hvm	Select 64-bit
	RStudio-1.1.383_R-3.4.2_Julia-0.6.0_CUDA-8_cuDNN-6_ubuntu-16.04-LTS-64bit - ami-3b0c205e Ready to run RStudio + Julia/Python server for statistical computation (www.louisaslett.com). Connect to instance public DNS in web browser (standard port 80), username rstudio and password rstudio Root device type: ebs Virtualization type: hvm	Select 64-bit
	OMPI_R_RStudioServer - ami-5de9cc38 [Copied ami-91b77587 from us-east-1] OMPI_R_RStudioServer Root device type: ebs Virtualization type: hvm	Select 64-bit
	FIN580-RStudio-Student - ami-70b99a15 FIN580-RStudio-Student Root device type: ebs Virtualization type: hvm	Select 64-bit
	microsoft-R-ubuntu - ami-b54464d0	Select

1. Choose AMI 2. Choose Instance Type 3. Configure Instance 4. Add Storage 5. Add Tags 6. Configure Security Group

Step 2: Choose an Instance Type

Amazon EC2 provides a wide selection of instance types optimized to fit different use cases. Instances are virtual servers that can run applications. They have varying combinations of CPU, memory, storage, and networking capacity, and give you the flexibility to choose the appropriate mix of resources for your applications. [Learn more](#) about instance types and how they can meet your computing needs.

Filter by: All instance types Current generation [Show/Hide Columns](#)

Currently selected: t2.micro (Variable ECUs, 1 vCPUs, 2.5 GHz, Intel Xeon Family, 1 GiB memory, EBS only)

	Family	Type	vCPUs	Memory (GiB)	Instance Storage (GB)	EBS-Optimized Available	Network Performance
<input type="checkbox"/>	General purpose	t2.nano	1	0.5	EBS only	-	Low to Mode
<input checked="" type="checkbox"/>	General purpose	t2.micro Free tier eligible	1	1	EBS only	-	Low to Mode
<input type="checkbox"/>	General purpose	t2.small	1	2	EBS only	-	Low to Mode
<input type="checkbox"/>	General purpose	t2.medium	2	4	EBS only	-	Low to Mode

[Cancel](#) [Previous](#) [Review and Launch](#) [Next: Configure Instance Details](#)

Step 3: Configure Instance Details

Configure the instance to suit your requirements. You can launch multiple instances from the same AMI, request Spot instances to take advantage of the lower pricing, assign an access management role to the instance, and more.

[Launch multiple instances](#)

Number of instances [Launch into Auto Scaling Group](#)

Purchasing option Request Spot instances

Network [Create new VPC](#) **Virtual Private Cloud allows users to isolate their environments**

Subnet [Create new subnet](#)

Auto-assign Public IP

IAM role [Create new IAM role](#) **For configuring user permissions**

Shutdown behavior

Enable termination protection Protect against accidental termination

Monitoring Enable CloudWatch detailed monitoring
Additional charges apply.

Tenancy **Users can launch instances on dedicated hardware**
Additional charges will apply for dedicated tenancy.

T2 Unlimited Enable
Additional charges may apply

[Cancel](#) [Previous](#) [Review and Launch](#) [Next: Add Storage](#)

Step 4: Add Storage

Your instance will be launched with the following storage device settings. You can attach additional EBS volumes and instance store volumes to your instance, or edit the settings of the root volume. You can also attach additional EBS volumes after launching an instance, but not instance store volumes. [Learn more](#) about storage options in Amazon EC2.

Volume Type	Device	Snapshot	Size (GiB)	Volume Type	IOPS	Throughput (MB/s)	Delete on Termination	Encrypted
Root	/dev/sda1	snap-03cebc94434eb0a6c	10	<input checked="" type="checkbox"/> General Purpose SSD (GP2) <input type="checkbox"/> Provisioned IOPS SSD (IO1) <input type="checkbox"/> Magnetic	100 / 3000	N/A	<input checked="" type="checkbox"/>	Not Encrypted

[Add New Volume](#)

Free tier eligible customers can get up to 30 GB of EBS General Purpose (SSD) or Magnetic storage. [Learn more](#) about free usage tier eligibility and usage restrictions.

[Cancel](#) [Previous](#) [Review and Launch](#) [Next: Add Tags](#)

aws Services Resource Groups Ohio Support

1. Choose AMI 2. Choose Instance Type 3. Configure Instance 4. Add Storage 5. Add Tags 6. Configure Security Group 7. Review

Step 5: Add Tags

A tag consists of a case-sensitive key-value pair. For example, you could define a tag with key = Name and value = Webservers. A copy of a tag can be applied to volumes, instances or both. Tags will be applied to all instances and volumes. [Learn more](#) about tagging your Amazon EC2 resources.

Key (127 characters maximum) Value (255 characters maximum) Instances Volumes

This resource currently has no tags

Choose the Add tag button or [click to add a Name tag](#). Make sure your [IAM policy](#) includes permissions to create tags.

Add Tag (Up to 50 tags maximum)

Cancel Previous Review and Launch Next: Configure Security Group

Feedback English (US) © 2008 - 2017, Amazon Web Services, Inc. or its affiliates. All rights reserved. Privacy Policy Terms of Use

aws Services Resource Groups Ohio Support

1. Choose AMI 2. Choose Instance Type 3. Configure Instance 4. Add Storage 5. Add Tags 6. Configure Security Group 7. Review

Step 6: Configure Security Group

A security group is a set of firewall rules that control the traffic for your instance. On this page, you can add rules to allow specific traffic to reach your instance. For example, if you want to set up a web server and allow Internet traffic to reach your instance, add rules that allow unrestricted access to the HTTP and HTTPS ports. You can create a new security group or select from an existing one below. [Learn more](#) about Amazon EC2 security groups.

Assign a security group: Create a new security group Select an existing security group

Security group name: packtbigdata 1. Set a name for the security group

Description: A Security Group for All TCP Traffic 2. Description

Type	Protocol	Port Range	Source	Description
All TCP	TCP	0 - 65535	Custom 0.0.0.0/0	e.g. SSH for Admin Desktop

Add Rule Change from default to All TCP

Warning
Rules with source of 0.0.0.0/0 allow all IP addresses to access your instance. We recommend setting security group rules to allow access from known IP addresses only.

Cancel Previous Review and Launch

aws
Services ▾ Resource Groups ▾
Ohio ▾ Support ▾

1. Choose AMI
2. Choose Instance Type
3. Configure Instance
4. Add Storage
5. Add Tags
6. Configure Security Group
7. Review

Step 7: Review Instance Launch

Please review your instance launch details. You can go back to edit changes for each section. Click **Launch** to assign a key pair to your instance and complete the launch process.

⚠ Improve your instances' security. Your security group, packtbigdata, is open to the world.

Your instances may be accessible from any IP address. We recommend that you update your security group rules to allow access from known IP addresses only. You can also open additional ports in your security group to facilitate access to the application or service you're running, e.g., HTTP (80) for web servers. [Edit security groups](#)

▼ AMI Details [Edit AMI](#)

RStudio-1.0.153_R-3.4.1_Julia-0.6.0_ubuntu-16.04-LTS-64bit - ami-12d1f277

Ready to run RStudio + (experimental) Julia server for statistical computation (www.louisaslett.com). Connect to instance public DNS in web browser (standard port 80), username rstudio and password rstudio

Root Device Type: ebs Virtualization type: hvm

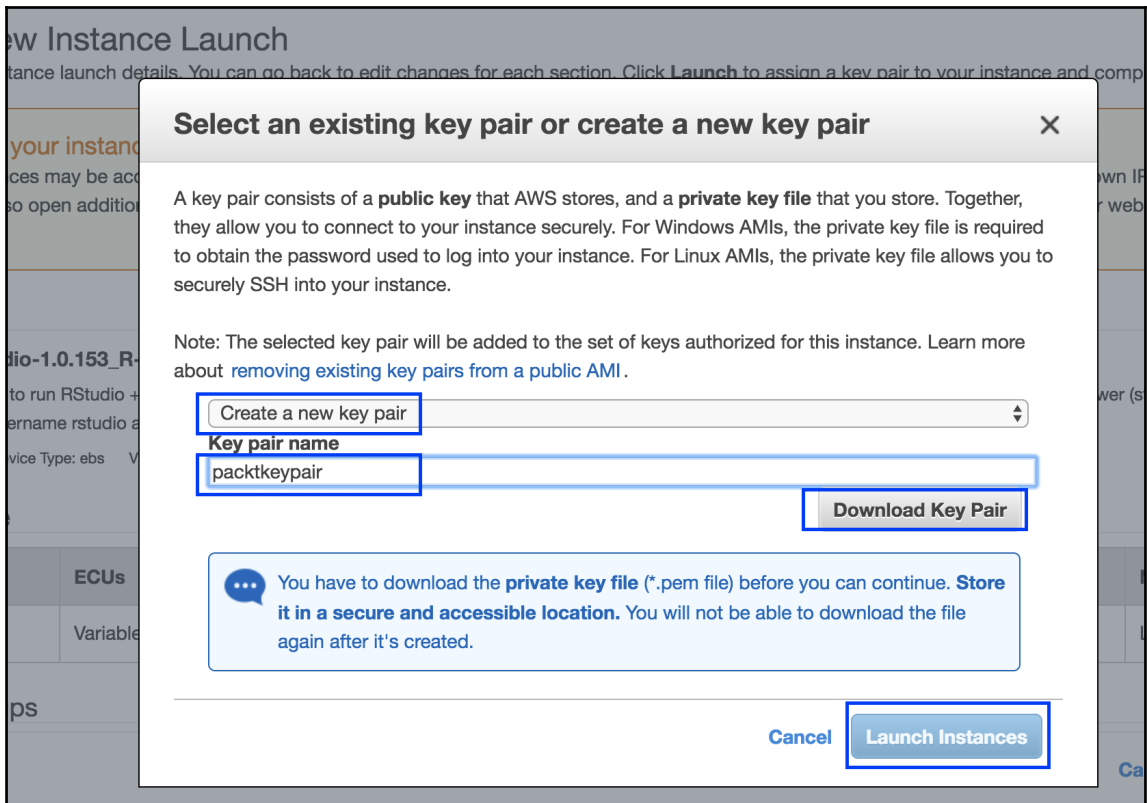
▼ Instance Type [Edit instance type](#)

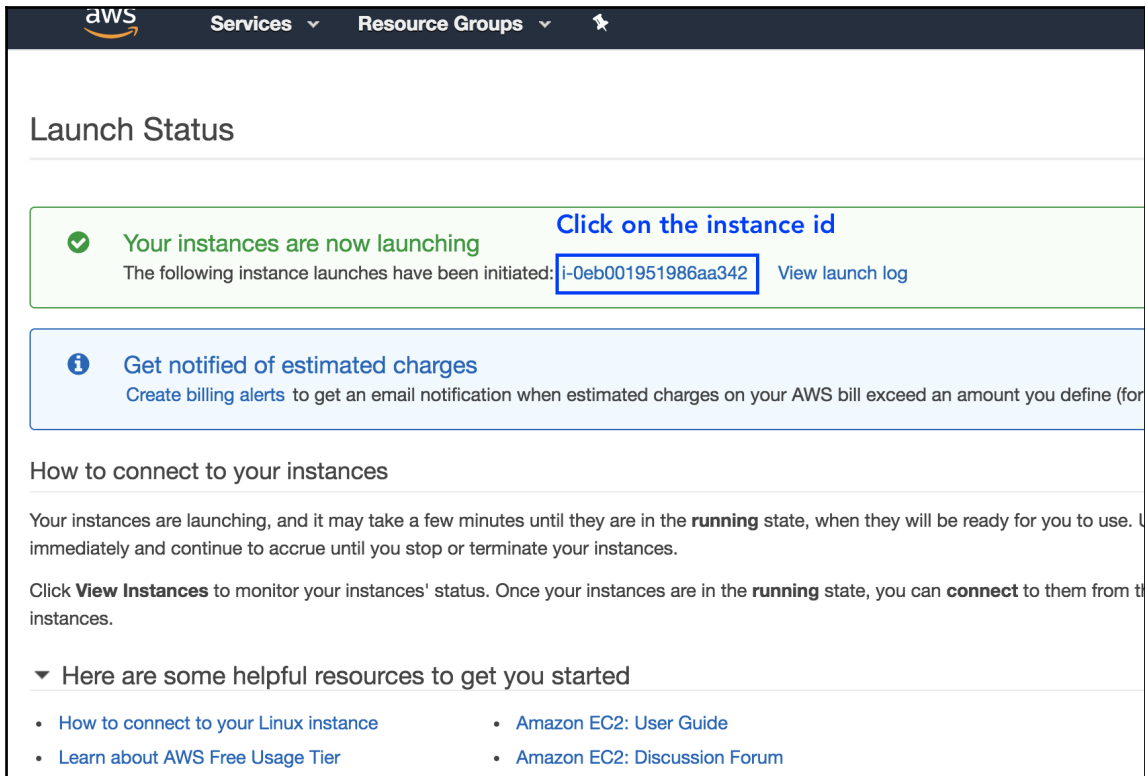
Instance Type	ECUs	vCPUs	Memory (GiB)	Instance Storage (GB)	EBS-Optimized Available	Network Performance
t2.micro	Variable	1	1	EBS only	-	Low to Moderate

▼ Security Groups [Edit security groups](#)

Cancel
Previous
Launch

Feedback
English (US)
© 2008 - 2017, Amazon Web Services, Inc. or its affiliates. All rights reserved. [Privacy Policy](#) [Terms of Use](#)





The screenshot shows the AWS Management Console interface. At the top, there is a navigation bar with the AWS logo, 'Services' dropdown, 'Resource Groups' dropdown, and a star icon. Below the navigation bar, the main content area is titled 'Launch Status'. There are two main sections: a green success message and a blue information message.

Launch Status

✓ Your instances are now launching [Click on the instance id](#)
The following instance launches have been initiated: [i-0eb001951986aa342](#) [View launch log](#)

i Get notified of estimated charges
[Create billing alerts](#) to get an email notification when estimated charges on your AWS bill exceed an amount you define (for

How to connect to your instances

Your instances are launching, and it may take a few minutes until they are in the **running** state, when they will be ready for you to use. You can connect to your instances immediately and continue to accrue until you stop or terminate your instances.

Click **View Instances** to monitor your instances' status. Once your instances are in the **running** state, you can **connect** to them from the **instances** page.

▼ Here are some helpful resources to get you started

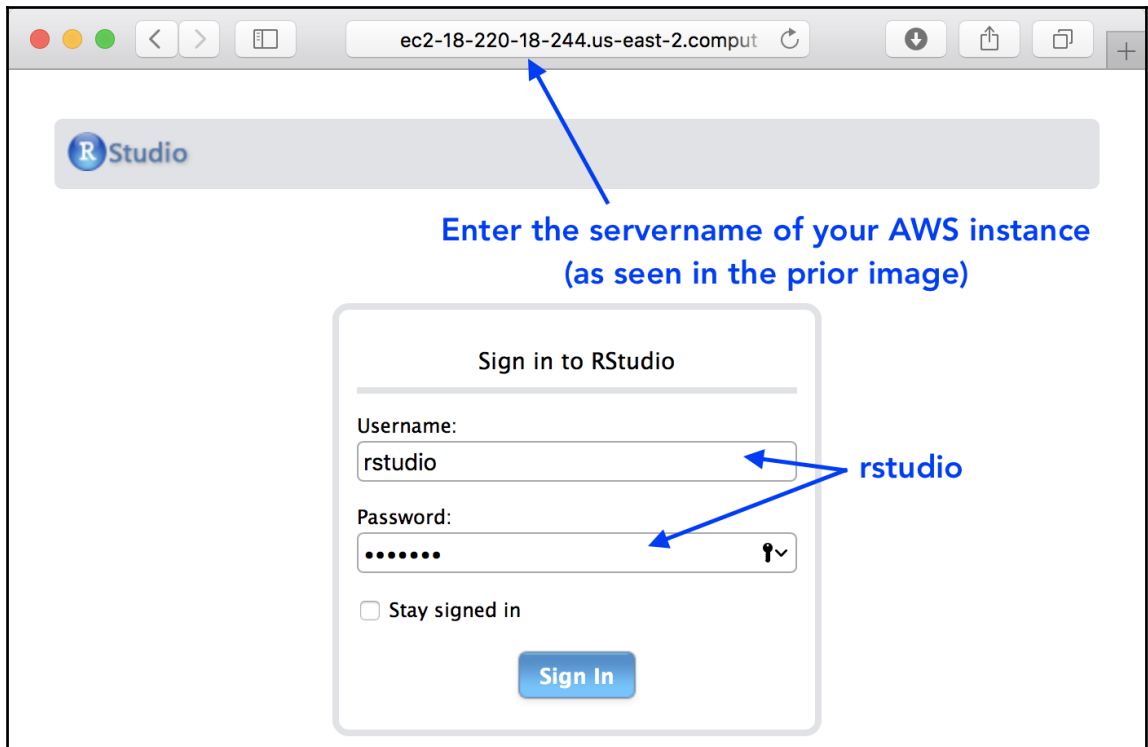
- [How to connect to your Linux instance](#)
- [Amazon EC2: User Guide](#)
- [Learn about AWS Free Usage Tier](#)
- [Amazon EC2: Discussion Forum](#)

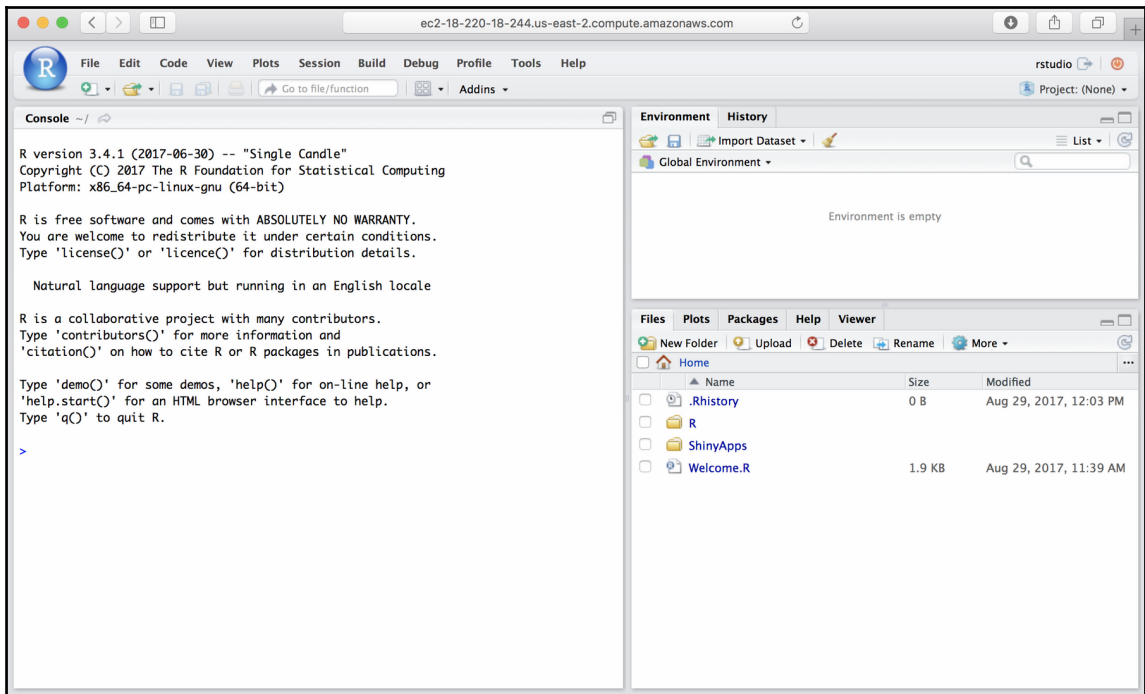
The screenshot displays the AWS Management Console interface for an EC2 instance. The top navigation bar includes the AWS logo, 'Services', 'Resource Groups', and a location dropdown set to 'Ohio'. The left-hand navigation pane lists various EC2-related services such as 'Instances', 'Images', and 'Network & Security'. The main content area shows a table with one instance listed, with its state highlighted as 'running'. Below the table, the instance details for 'i-0eb001951986aa342' are shown, including its public DNS name and instance type. A 'Copy this server name' button is visible next to the public DNS field.

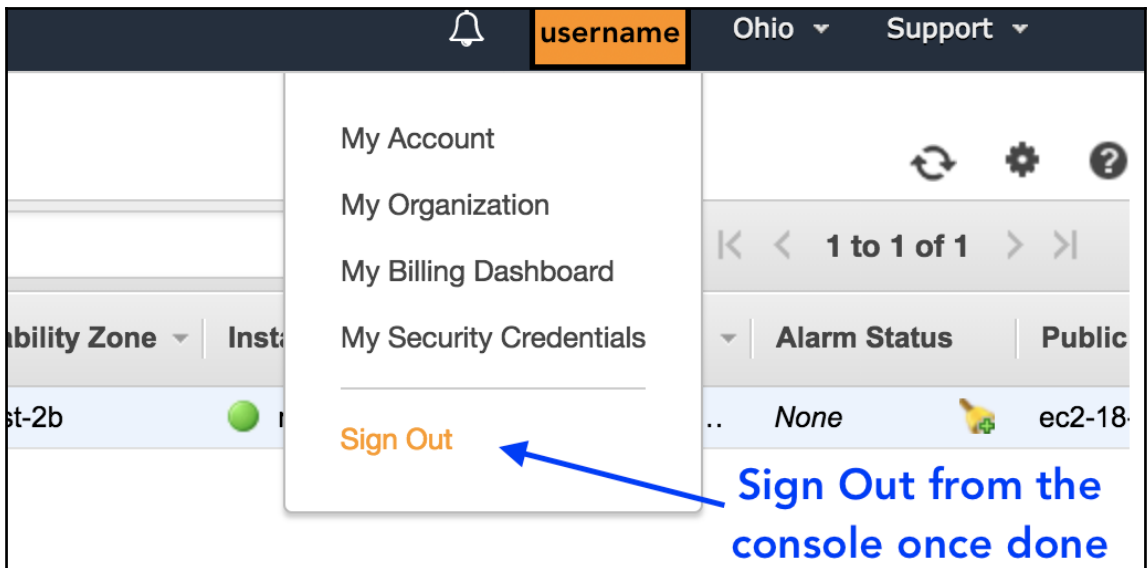
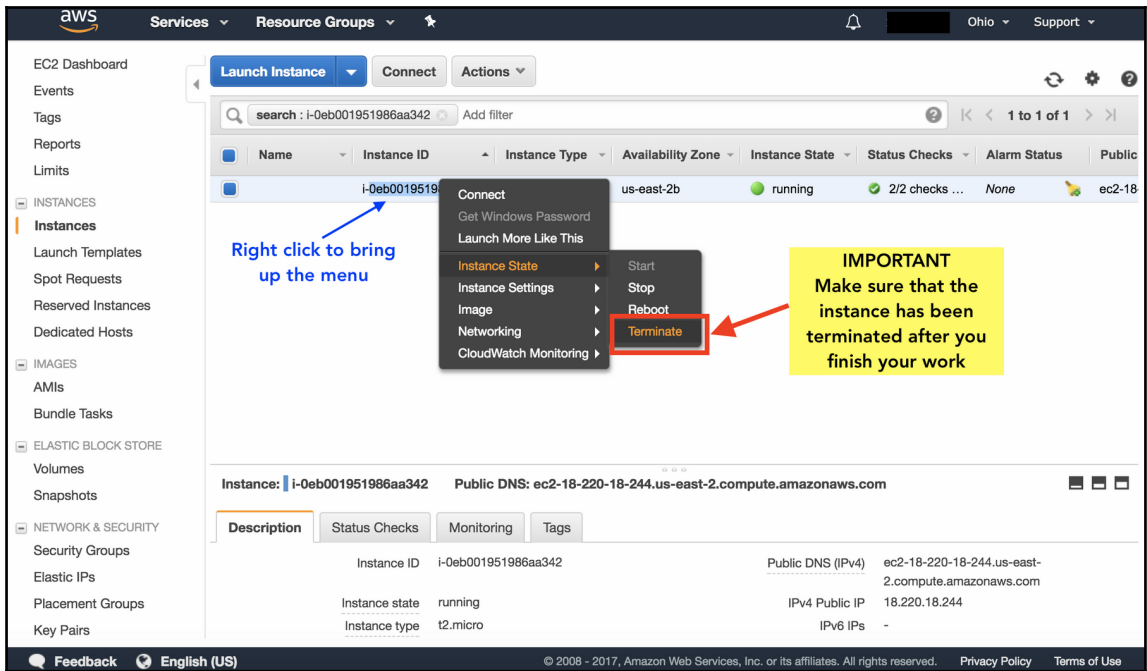
Name	Instance ID	Instance Type	Availability Zone	Instance State	Status Checks	Alarm Status	Public
	i-0eb001951986aa342	t2.micro	us-east-2b	running	Initializing	None	ec2-18

Instance: **i-0eb001951986aa342** Public DNS: **ec2-18-220-18-244.us-east-2.compute.amazonaws.com**

Description		Status Checks	Monitoring	Tags
Instance ID	i-0eb001951986aa342	Public DNS (IPv4)	ec2-18-220-18-244.us-east-2.compute.amazonaws.com	
Instance state	running	IPv4 Public IP	18.220.18.244	
Instance type	t2.micro	IPv6 IPs	-	







Microsoft Azure

SALES 1-800-867-1389 MY ACCOUNT PORTAL Search

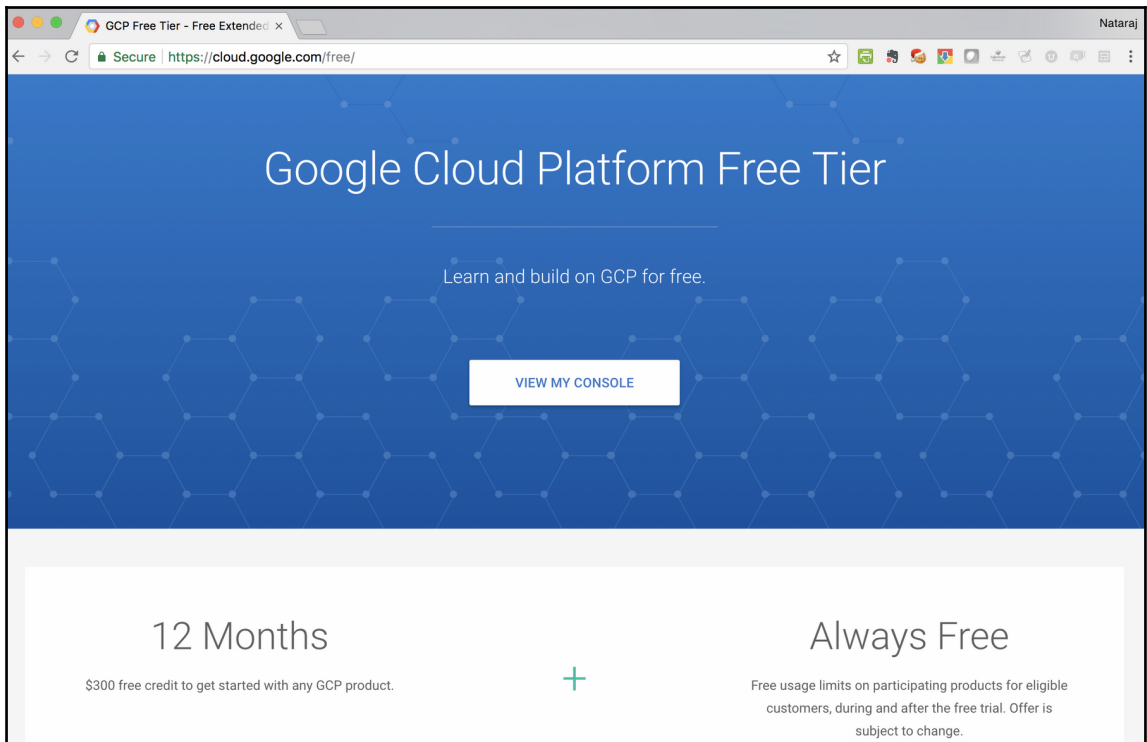
Why Azure Products Solutions Documentation Pricing Training Marketplace Partners Blog Resources Support

NEW Now get 12 months of free access to popular products.

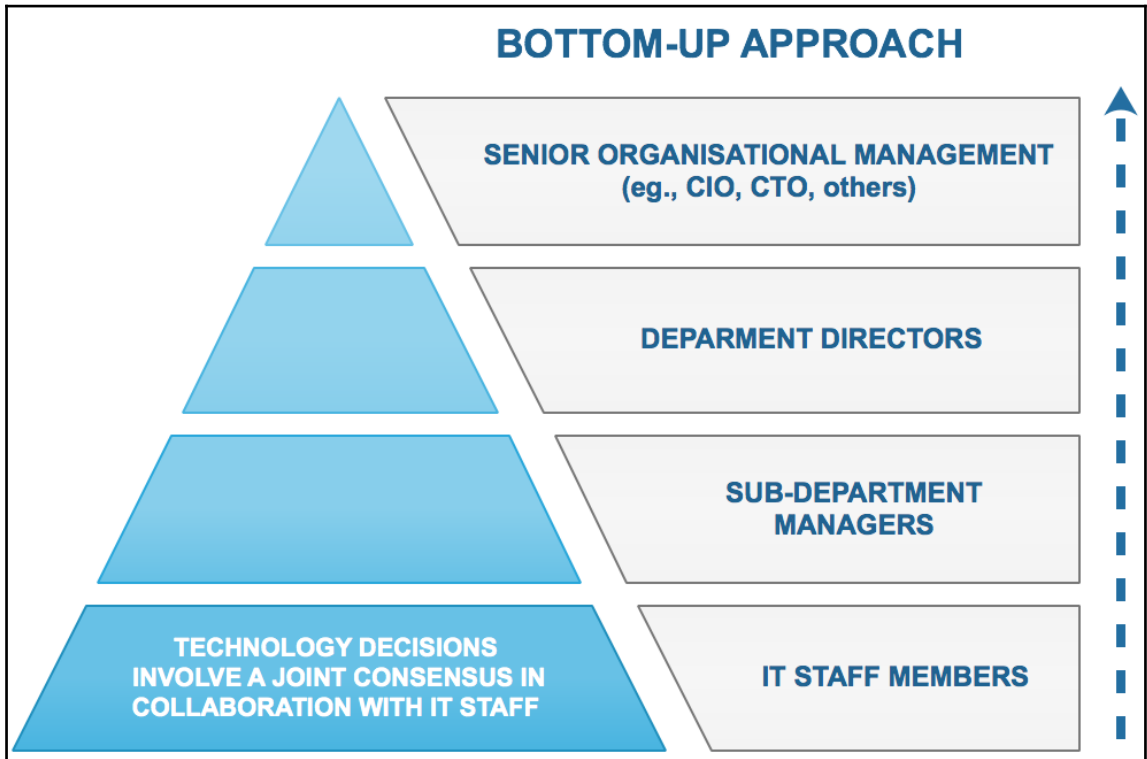
Create your Azure free account today

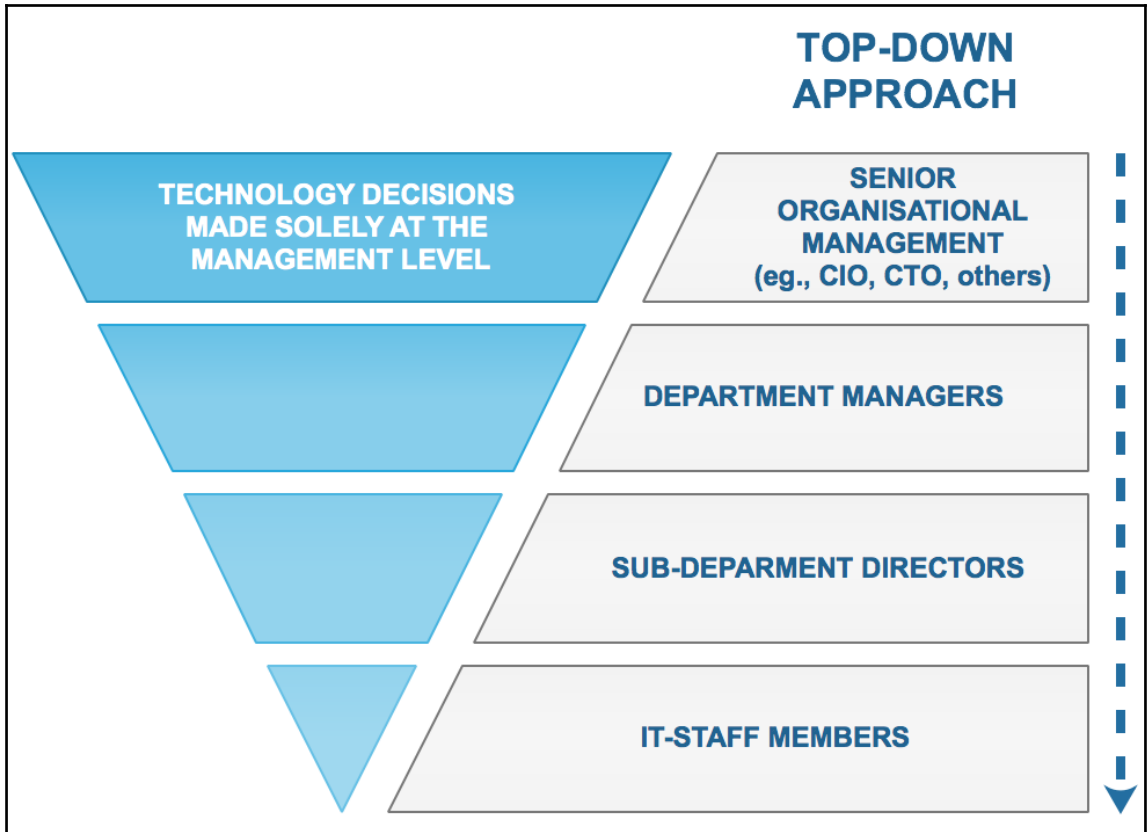
- Get started with a \$200 credit**
Start free with \$200 in credit to use on any Azure products for 30 days.
- Keep going with free products**
Build your next great idea with free access to our most popular products for 12 months and to more than 25 always free products.
- Pay nothing until you choose**
We use your credit card information for identity verification, but you will not be charged until you choose to upgrade.

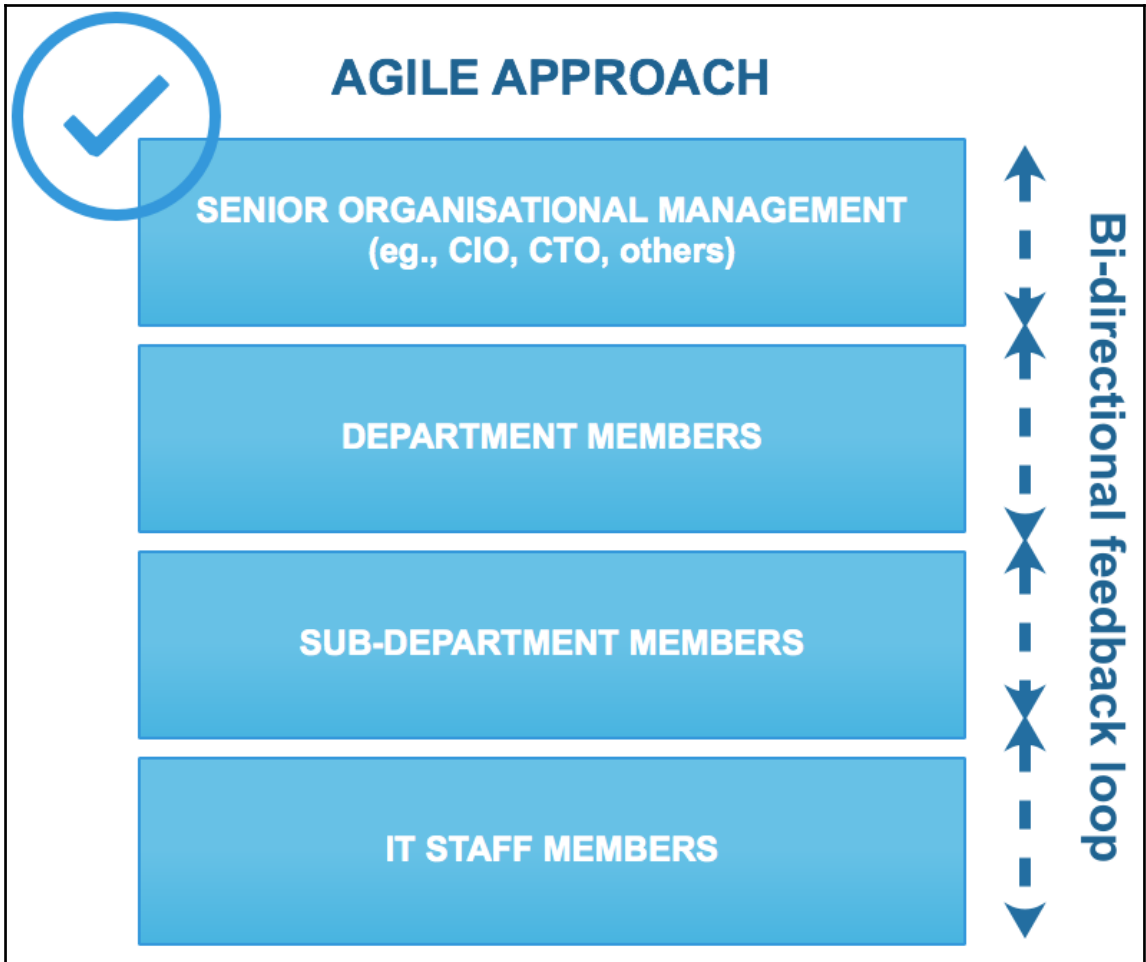
[Start free >](#)



Chapter 10: Closing Thoughts on Big Data







crunchbase

[Products](#)
[Resources](#)
[News](#)
[About](#)

[LOG IN](#)
[REGISTER](#)

Introducing Crunchbase Pro [LEARN MORE](#)

Discover innovative companies and the people behind them

SEARCH

- [Companies](#)
- [People](#)
- [Investors](#)
- [Funding Rounds](#)
- [Acquisitions](#)
- [Schools](#)
- [Events](#)
- [My Searches](#)
- [My Lists](#)
- [Featured](#)
- [Add New Profile](#)

[About](#) • [Terms](#) • [Careers](#) • [Sitemap](#)
 Editorial Partners: [Oath Tech](#)
 © 2017 Crunchbase Inc.
 All rights reserved.

Latest insights and analysis

December 21, 2017

\$4 Billion More For Ridesharing, Because Why Not

Alex Wilhelm

Morning Report: Didi raises another chunk of cash. Is \$4 billion a lot?

[Continue Reading](#)

December 20, 2017

Looking Back At Softbank's Big Year

Joanna Glasner

In an effort to pinpoint the startups spaces that are see the greatest impact from SoftBank's largess, Crunchbase drilled into the activities of the firm's \$100 billion Vision

[Continue Reading](#)

December 20, 2017

Today In Uber: New COO, EU Troubles, Rivals On The Move

Alex Wilhelm

Morning Report: Here's the pace of news at Uber—a firm in transition

Activity this week

373

\$ Funding rounds announced

16b

\$ Total Fundings

173

➔ Acquisitions recorded

87.5b

➔ Total Acquisitions Amount

Featured searches and lists

Name	Entries
Initial Coin Offerings	204
Companies that have recently changed CEOs	132
Acquisitions of the past week	172
Coffee startups	40
International Fintech companies with > 5M fundi...	383